Reverse stress testing for supply chains

Busra Gultekin[1], Frans de Ruiter[1], Meike Reusken[1], and Sander de Leeuw[1]

[1]Operations Research and Logistics Group, Wageningen University, Wageningen 6706 KN, The Netherlands

{busra.b.gultekin, frans.deruiter, meike.reusken, sander.deleeuw}@wur.nl

Abstract

This study introduces reverse stress testing for supply chains, designed to identify the minimal deviations from normal operations that would drive supply chains to a predefined performance failure. First, we present a framework for reverse stress testing with the purpose of assessing supply chain vulnerabilities. The framework involves six steps: selecting risk variables, defining baselines, formalizing the policy, formalizing key performance indicators (KPIs), determining critical thresholds, and identifying the critical scenario. Building on this framework, we propose an optimization model that identifies the critical scenario by uncovering the joint risk variable deviations that stress the supply chain beyond critical KPI thresholds. We also show that the optimization model has desirable theoretical properties, demonstrating that the model is computationally tractable, that a stress scenario always exists under mild assumptions, and that its output establishes a robustness guarantee for the policy. Two numerical examples demonstrate the framework's applicability: a serial supply chain model and a service guarantee model. Results show that the framework effectively uncovers where disruptions become most harmful.

Keywords: Supply chain resilience, stress testing, optimization, risk identification.

1 Introduction

Risk management has become more critical than ever, especially in the face of unforeseen events that can severely disrupt supply chains. Natural disasters, pandemics, and geopolitical conflicts are notoriously difficult to foresee, making it hard for businesses to anticipate disruptive events and mitigate their impact (Hosseini et al., 2019). Events such as the COVID-19 pandemic, the war in Ukraine, and past crises like the 2011 Japan earthquake and Thailand floods have highlighted the vulnerabilities in supply chains, causing resource shortages, production delays, and rising costs (Guan et al., 2020; Allam et al., 2022; Hosseini et al., 2019; Haraguchi and Lall, 2015). Besides these black swan events, there are several smaller uncertainties that together in an unlucky combination or compounding effect cause stress in the supply chain. For example, all uncertainties in the infant dairy formula supply chain (Al-Khatib et al., 2024) or the recent stress caused in automotive production caused by supply issues in a single semiconductor company that produces standard, so called mature node chips (Reuters, 2026). These disruptions have reinforced the need for organizations to develop resilient supply chain strategies that enhance adaptability and minimize risks. As a result, companies that proactively prepare for unexpected shocks gain a competitive advantage ensuring long-term stability and operational continuity in an increasingly volatile environment (Chopra and Sodhi, 2014; Wicaksana et al., 2022).

To mitigate risk, supply chain management has often relied on traditional risk management approaches. Such approaches start with making an inventory of risks and the establishment of the likelihood and impact of those risks. In these traditional approaches, once likelihoods and impact of these risks have been established, the next step is to identify possible scenarios to mitigate risks and then to simulate the impact of those scenarios. We refer to this as *forward stress testing*. Even though it is widely used, forward stress testing is bound by a tendency to underestimate low-probability, high-impact events or joint, compound small effects creating significant blind spots in risk assessment (Chopra and Sodhi, 2014; Gunasekaran et al., 2015). This shortcoming exposes supply chains to unexpected disruptions and raises doubts about whether these methods effectively support realistic mitigation strategies (Pettit et al., 2010). Risk mitigation policies, whether in corporate or governmental settings, often are difficult to establish due to the many uncertainties and unknown probabilities involved (Hollis and Ekengren, 2025). This underscores the need for more adaptive, probability agnostic and comprehensive approaches to managing supply chain risks amid rising uncertainties.

In this paper, we introduce and detail the concept of *reverse stress testing* of supply chains, which is inspired by research in financial economics. We develop a formal framework for reverse stress testing and demonstrate how using this framework critical scenarios can be identified starting from critical business outcomes. Unlike forward stress testing, which starts with predefined risk scenarios and evaluates their impact on business outcomes, reverse stress testing begins by identifying the most critical business outcomes such as severe shortages, prolonged delivery delays, or complete operational breakdowns—and then traces back to the conditions that could cause these critical business outcomes or failures. By identifying the critical failures first, rather than relying on probability-based assessments of failure risks, and then working backwards from critical failures to identify which events are required to trigger such a failure, reverse stress testing can provide companies with actionable insights. These insights enable the development of targeted mitigation strategies that enhance supply chain resilience.

In more detail, the contributions of this paper are the following

1. We introduce a six-step reverse stress testing framework for supply chains, outlining the key elements and methodology. This offers insights for evaluating policies and, if needed, developing mitigation strategies.

2. We introduce an optimization model for the last step of this framework, which identifies the critical scenario leading to supply chain performance failure. Applying simple techniques to generate a variety of multiple scenarios, we further provide a broader perspective on potential systemic vulnerabilities.

3. We prove three theoretical properties of our reverse stress test model that support its effectiveness and efficiency. First, we show that our model is tractable. Specifically, depending on the chosen distance measure, KPI functions, and policies, the model can be formulated as an (integer) linear, piecewise-linear, or convex quadratic program, which can be efficiently solved by existing optimization solvers. Second, we prove the guaranteed existence of a stress scenario under mild assumptions. Third, we provide a formal guarantee of policy robustness by demonstrating that all scenarios within a specified distance from the baseline satisfy the KPI thresholds.

4. We demonstrate the practical application of the supply chain framework and optimization model by applying it to two supply chain networks: a serial supply chain example and a service guarantee supply chain model based on Simchi-Levi et al. (2018).

The remainder of the paper is organized as follows: Section 2 covers the relevant literature on stress testing in supply chains. We present our reverse stress testing framework for supply chains in Section 3, including a mathematical optimization model for executing the sixth step of the framework. We explain the theoretical properties of this model in Section 4. The numerical results are given and analyzed in Section 5. Concluding remarks are provided in Section 6.

2 Literature

Reverse stress testing in supply chains, as introduced in this paper, relates to three areas of research: forward stress testing in supply chains, reverse stress testing in the financial sector, and several aspects of optimization theory, including inverse, counterfactual, and robust optimization methods. The following subsections review these scientific areas.

2.1 Stress testing in supply chains

Building on insights from a recent review (Ivanov, 2025) and a special issue (Dolgui et al., 2025), this section discusses the application of stress testing in supply chains. Stress testing from the literature, referred to as forward stress testing in this study, is a possible risk management method to identify vulnerabilities and improve supply chain resilience in various scenarios (Khakifirooz et al., 2025). The main elements of stress testing are a predefined scenario, a stress testing model, and performance impact analysis (Dolgui et al., 2025).

The most frequently examined shocks involve disruptions to suppliers, demand, and capacity, as well as ripple and bullwhip effects. To evaluate supply chain resilience under such disruptions, the literature employs a variety of performance indicators. The most common measures are service level and cost, while other studies utilize revenue, lead time, inventory levels, recovery measures such as Time-to-Recover (TTR) and Time-to-Survive (TTS), utility, and various specific resilience and risk metrics. To capture these behaviors and assess the resulting performance impacts, researchers predominantly employ simulation-based methods (such as discrete-event simulation), alongside robust and stochastic optimization. Furthermore, various analytical methods—including Bayesian networks, Markov chains, network theory, machine learning, and the possible maximum loss method—are widely applied. Several studies comprehensively explore these disruptions, performance indicators, and methodologies (e.g., Brintrup et al., 2020; Bruetzel et al., 2025; Diem et al., 2025; Dolgui et al., 2020; Gao et al., 2019; Hosseini et al., 2020; Hosseini and Ivanov, 2022; Ivanov, 2017, 2020, 2024; Kinra et al., 2020; Li and Zobel, 2020; Li et al., 2021; Liu et al., 2021, 2024; Sawik, 2022; Simchi-Levi et al., 2015).

These studies demonstrate that supply chain stress tests are designed to investigate disruption scenarios, analyze resulting supply chain behavior through modeling approaches, and evaluate performance impacts. As such, both forward and our reverse stress testing aim to assess supply chain vulnerabilities and inform resilience strategies. The key distinction between forward and reverse stress tests lies in their methodology. Forward variants assess how specific predefined scenarios affect supply chain performance, whereas reverse stress testing explores the scenario that could push the supply chain to critical business outcomes. Unlike forward stress testing, where scenarios must be explicitly stated in advance, it directly identifies such a scenario leading to critical business impact.

2.2 Reverse stress testing in the financial sector

The idea of reverse stress testing is originally introduced within the financial sector as a means to identify risks and define mitigation measures. Following the financial crises in 2008, the limitations of forward stress tests became apparent, particularly in their inability to capture extreme yet plausible risks (Breuer and Csiszár, 2013). As a result, major regulators began to emphasize the importance of conducting reverse stress tests (Basel Committee on Banking Supervision, 2009; Committee of European Banking Supervisors, 2009; Financial Services Authority, 2008). In the last decade, financial institutions in the United Kingdom and the European Union are now required to perform reverse stress tests using quantitative methods (Grigat and Caccioli, 2017).

In the financial sector, Grundke (2011) introduced a quantitative reverse stress testing model that identifies the specific level of systemic risk factors where the combined impact of expected losses and extreme potential losses (value at risk) consumes the institution's available capital buffer.

The most commonly examined systemic risk factors are credit risk (Grundke, 2011; Grundke and Pliszka, 2018; Grigat and Caccioli, 2017; Albanese et al., 2023; Liu et al., 2023), interest rate risk (Grundke, 2011; Grundke and Pliszka, 2018; Montesi et al., 2020), and market risk (Albanese et al., 2023). These studies apply various quantitative methods to identify stress scenarios and uncover key insights about systemic risks. Simulation techniques such as Monte Carlo methods, together with parameter search techniques like grid search, are commonly used to identify the probable stress scenarios and their associated likelihoods (Grundke, 2011; Grundke and Pliszka, 2018; Albanese et al., 2023). Optimization methods, including both deterministic models (Grigat and Caccioli, 2017; Mitic, 2021) and stochastic algorithms like simulated annealing (Montesi et al., 2020), have been used to identify minimal shocks or conditions leading to critical capital breaches. Additionally, statistical tools like principal component analysis together with orthogonalization and likelihood maximization enable the generation of diverse and plausible stress scenarios (Kopeliovich et al., 2015). Collectively, these methodologies provide a deeper understanding of the most probable scenarios, supporting regulators and financial institutions in enhancing their resilience against systemic shocks.

Despite the attention from regulators, reverse stress testing still lacks clear guidelines and remains scarce even within the financial sector (Mitic, 2021; Albanese et al., 2023) let alone beyond the financial area. Building on this gap, we develop a framework for supply chains that not only introduces the methodology to supply chains, but also establishes the structured guiding framework. We do so by integrating context-specific risk variables, operational policies, and supply chain performance KPIs to identify the minimal shock that leads to critical failure. Thus, it provides actionable insights tailored for supply chain resilience rather than financial solvency.

2.3 Optimization theory: Inverse optimization, counterfactual explanations, and robust optimization

Our reverse stress testing framework uses mathematical optimization to identify the critical scenario that could cause a policy to fail. Traditionally, optimization techniques have been used to construct optimal solutions or best-performing policies within specific classes, such as order-up-to policies and perform sensitivity analysis examining how these solutions change as input parameters vary. This perspective differs from reverse stress testing, which assesses how a given solution or policy performs when exposed to adverse scenarios without policy adjustments in hindsight. Nevertheless, several optimization techniques are conceptually related to reverse stress testing, particularly inverse optimization, counterfactual explanations, and robust optimization. In the following, we briefly review each of these techniques and discuss their similarities and differences with reverse stress testing.

1. **Inverse optimization and counterfactual explanations.** Inverse optimization estimates the parameters of an optimization problem that make observed decisions optimal or near-optimal, and counterfactual explanations identifies minimal changes to input parameters that would lead to a particular outcome. Similar to reverse stress testing, these approaches examine the problem in reverse, tracing outcomes back to the parameters or conditions that influence them.

   Inverse optimization estimates the parameters of a forward optimization problem that justify observed decisions, using a prior belief or estimate as a reference. In contrast, while reverse stress testing similarly uses a baseline scenario as a reference, it estimates the deviations in variables that result in the violation of performance constraints. While inverse optimization relies on optimality conditions or data-driven formulations to define parameter sets, our reverse stress testing framework explicitly formulates critical thresholds and policies to specify stress conditions. For further explanation on inverse optimization, see Chan and Kaw (2020); Gorissen et al. (2025), and Chan et al. (2025).

   Counterfactual explanations are closely related to inverse optimization but differ in their input–output relationships, solution specificity, and computational complexity. In inverse optimization, the desired optimal solution is usually predefined, whereas in counterfactual explanations, the outcome is not fixed in advance. Instead, it is defined implicitly through constraints, and the optimization seeks minimal changes in the input parameters to achieve that outcome. This makes counterfactual explanations computationally more challenging than inverse optimization. In reverse stress testing, the desired outcome or scenarios are not prespecified either, but are determined through KPIs and their corresponding critical thresholds. Unlike counterfactual explanations, where parameter changes are usually limited to predefined intervals, reverse stress testing does not impose such restrictions; instead, it explores the entire predefined feasible region to identify the critical scenario. For an in-depth introduction on the recent research on counterfactual explanations, see Kurtz et al. (2025); Korikov et al. (2021); Korikov and Beck (2021, 2023).

2. **Robust optimization.** Robust optimization is a widely studied method for decision-making under uncertainty that became widespread through the work of Ben-Tal and Nemirovski (2002); Ben-Tal et al. (2004); Bertsimas and Sim (2004); Ben-Tal et al. (2015), among others. It defines an uncertainty set to specify the values of uncertain parameters for which the decision must remain feasible. The model is formulated to ensure that constraints are satisfied for every value within this set (Ben-Tal and Nemirovski, 1998). In practice, a key challenge is determining an appropriate form for the uncertainty set. It must align with available data, be statistically meaningful, and ensure that the resulting problem remains tractable. For more background on robust optimization, we refer to the books by Ben-Tal et al. (2009b) and Bertsimas and den Hertog (2022).

   Reverse stress testing and robust optimization can work complementarily to enhance resilience in decision-making under uncertainty. Robust optimization could be used if no policy is implemented yet, or to find alternative policies that perform better under reverse stress tests. Importantly, as shown in Section 4.3, the reverse stress testing model provides a guarantee that a given policy (robust or not) remains feasible for all scenarios within a specified distance from the baseline. Furthermore, to prevent overly conservative solutions arising from large uncertainty sets, robust optimization involves trade-offs in the size of the set, which may exclude stress scenarios that were not imagined beforehand or lead to excessively pessimistic solutions. In contrast, our reverse stress testing framework does not require an explicitly defined uncertainty set, allowing it to directly assess the resilience of a given solution or policy by revealing the conditions under which it may fail.

3 A reverse stress testing framework and model for supply chains

The reverse stress testing framework consists of six steps: (1) selecting risk variables that contain potential sources of uncertainty, (2) defining baselines, which represents the normal operating supply chain conditions, averages or management targets, (3) formalizing the policy, i.e. the set of rules that govern operations, (4) formalizing KPIs that represent the financial or operational health in the supply chain, (5) determining critical thresholds at which disruptions transition from manageable inconveniences to dangerously severe operational risks, and (6) identifying the critical scenario that causes one or several KPIs to fall below their critical threshold, thus marking the first failure point at which the supply chain no longer meets requirements. For this sixth step, this section presents a reverse stress testing optimization model to identify the critical scenario. A more detailed description of each of the steps can be found in Appendix A, and the flow and connection of this six-step reverse stress testing framework is shown in Figure 1.

The critical scenario identified through this framework represents the specific set of conditions that push the supply chain beyond its performance requirements. By comparing this scenario with the current operating conditions, the supply chain's vulnerabilities and weak points can be uncovered. While the framework itself ends at the identification of this critical scenario, such results can serve as input for managerial evaluation and decision-making, for instance, to explore policy adjustments or adaptive measures to the supply chain.

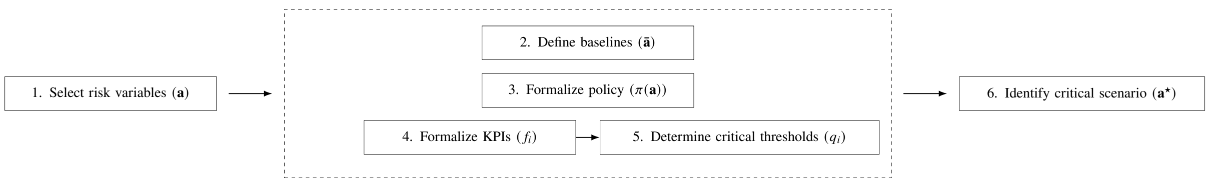


Figure 1: The reverse stress testing framework for supply chains. Steps 2, 3, and 4–5 can be carried out in parallel.

To identify the critical scenario as part of the final step of the framework, we now introduce a reverse stress testing optimization model for supply chains. Let $\mathbf{a} \in S$ represent risk variables such as demand, lead time, or costs, where boldface letters denote vectors throughout this paper. The scenario set $S \subseteq \mathbb{R}^n$ represents all possible realizations of risk variables in the supply chain without limitations such as boundedness. For example, when customer demand is the sole risk variable and no returns are allowed (e.g., perishable foods or other consumables), S equals $\mathbb{R}^n_+$.

The baseline scenario, which represents the normal operating conditions of these variables, is denoted by input parameter $\bar{\mathbf{a}} \in \mathbb{R}^n$. Depending on the nature of each risk variable, this baseline is derived from historical averages, forecasts, planned values or managerial targets.

Let $\mathcal{X}$ denote the set of all feasible operational actions, such as allowable production or replenishment levels. A policy defines the specific action $\pi(\mathbf{a}) \in \mathcal{X}$ for any observed realization of the risk variables $\mathbf{a} \in S$. To illustrate, if $\mathbf{a}$ denotes the customer demand, the policy $\pi(\mathbf{a})$ determines the operational action, such as the specific replenishment quantity to order in response to that demand level. Given a realization of the risk variable $\mathbf{a}$, the performance of the policy $\pi(\mathbf{a})$ is evaluated on various aspects of operational effectiveness using a set of m KPI functions: $f_i : \mathbb{R}^n \times \mathcal{X} \to \mathbb{R}, i = 1, \ldots, m$, where $f_i(\mathbf{a}, \pi(\mathbf{a}))$ depends on both the risk variables $\mathbf{a}$ and the policy $\pi(\mathbf{a})$. In other words, the KPI reflects not only the risk variables in the supply chain (e.g., demand, lead time, costs) but also how the policy reacts to those conditions. Using the same example, if $\mathbf{a}$ represents customer demand and $\pi(\mathbf{a})$ specifies how much to order, the KPI function could measure service level or unmet demand, which is affected both by the demand values and by the ordering decision.

For each KPI f_i, a critical threshold $q_i \in \mathbb{R}$ is defined. A failure occurs when at least one KPI falls below or equals its critical level, i.e., $f_i(\mathbf{a}, \pi(\mathbf{a})) \leq q_i$. The set of stress scenarios that lead to a failure of $\pi(\mathbf{a})$ is therefore:

$$F = \{\mathbf{a} \in S \,|\exists i\ f_i(\mathbf{a}, \pi(\mathbf{a})) \leq q_i,\ i = 1, \ldots, m\}. \tag{1}$$

We are interested in identifying the critical scenario closest to the baseline $\bar{\mathbf{a}}$ that causes the supply chain to violate its performance requirements. The notion of closeness is quantified by a distance measure $g : S \times S \to \mathbb{R}^+$, such as an ℓ_p norm $g(\mathbf{a}, \bar{\mathbf{a}}) = \|\mathbf{a} - \bar{\mathbf{a}}\|_p$. When multiple risk variables with different units or scales are considered, distance measures whose values depend on the variable scale (for example ℓ_p norm) may be dominated by variables with larger numerical ranges, making input normalization necessary.

The reverse stress testing model, identifying the critical scenario in step 6 of the framework in Figure 1, is formulated as:

$$\min_{\mathbf{a} \in F} g(\mathbf{a}, \bar{\mathbf{a}}), \tag{2}$$

where $\mathbf{a}$ are treated as optimization variables. The minimizer $\mathbf{a}^\star$ to (2) is the critical scenario, which represents the smallest deviation from $\bar{\mathbf{a}}$ that pushes the supply chain with policy $\pi(\mathbf{a})$ its critical KPI values. Unlike models that depend on strong assumptions, such as known probability distributions or fixed uncertainty bounds (Simchi-Levi et al., 2018), we impose only minimal structure on the scenario set S to keep the model general and adaptable to various contexts.

The resulting optimal solution from (2) provides in principle a single critical scenario. However, there could be multiple stress scenarios at the same or very similar distance from the baseline. This is a general phenomenon for optimization models where there can be multiple (near) optimal solutions. To obtain a variety of stress scenarios we propose two methods. Either one can vary the distance measure, e.g. by using different $p-$norms, or one could perform a random search in the neighborhood of the optimal solution. The second approach can be achieved as follows (Berntsen and Trutnevyte, 2017). Suppose that the initial optimization run of (2) yields an objective value of Γ. Uniformly draw a vector $c \in \mathbb{R}^n$ and solve for a given (small) $\epsilon > 0$:

$$\min_{\mathbf{a} \in F,\ g(\mathbf{a}, \bar{\mathbf{a}}) \leq \Gamma(1+\epsilon)} c^\top a. \tag{3}$$

This can be repeated several times to find alternative solutions. Note that in practice, not each run will result in a distinct stress scenario, as many objective directions c have the same optimal solution for $a \in F$. These two steps are also demonstrated via our numerical example in Section 5.1.4.

4 Theoretical properties

The reverse stress testing model (2) has several desirable properties. First, we demonstrate its tractability, showing the model is linear or convex for commonly used distance measures and KPI functions. Second, we establish its effectiveness, showing that under mild assumptions, it is guaranteed to identify a critical scenario. Finally, we highlight its connection to robust optimization, as the reverse stress testing model provides a robustness guarantee for the policy up to the critical scenario distance.

4.1 Tractability

The complexity of an optimization problem is primarily defined by the structure of the objective function and constraints. In our reverse stress testing model, complexity is thus driven by the distance function $g(\cdot)$ in the objective and the KPI functions $f_i(\cdot)$ defining the feasible region, so we next analyze their characteristics.

A distance function quantifies the dissimilarity between two points, such as the scalars a and $\bar{a}$. If this function satisfies the properties of non-negativity (i.e., $g(a,\bar{a}) \geq 0$), identity of indiscernibles (i.e., $g(a,\bar{a}) = 0 \iff a = \bar{a}$), and symmetry (i.e., $g(a,\bar{a}) = g(\bar{a},a)$), it is called a *distance measure*. If one or more of these properties are not satisfied, it is called a *similarity measure*. A distance measure that also satisfies the triangle inequality (i.e., $g(a,\bar{a}) \leq g(a,b) + g(\bar{a},b)$) is called a *distance metric* (Weller-Fahy et al., 2014).

Since the properties of a distance measure are sensible for supply chains, we focus on this definition for our distance function.

Note that distance measures are not necessarily convex. For example, the simple discrete metric

$$g(a,\bar{a}) = \begin{cases} 0 & \text{if } a = \bar{a}, \\ 1 & \text{if } a \neq \bar{a}, \end{cases}$$

satisfy non-negativity, identity of indiscernibles, and symmetry, but are not convex. However, realistic distance measures used in supply chains are typically convex. To support this, Table 1 lists common distance measures for supply chains. The first column shows the name and common abbreviation, the second the mathematical formula, and the last the resulting complexity after applying standard reformulations such as an epigraph formulation and linearization using auxiliary variables. This shows that, after reformulation, typical distance measures are linear or convex.

Distance measure	$g(\mathbf{a},\bar{\mathbf{a}})$	**Complexity**
Manhattan (ℓ_1)	$\sum_{j=1}^{n} \lvert a_j - \bar{a}_j \rvert$	Linear
Euclidean (ℓ_2)	$\sqrt{\sum_{i=1}^{n} (a_j - \bar{a}_j)^2}$	Convex quadratic
General (ℓ_p, $p > 2$)	$\left(\sum_{j=1}^{n} \lvert a_j - \bar{a}_j \rvert^p\right)^{\frac{1}{p}}$	Convex nonlinear
Max norm (ℓ_∞)	$\max_{1 \leq j \leq n} \lvert a_j - \bar{a}_j \rvert$	Linear
Mahalanobis	$\sqrt{\sum_{j=1}^{n} \sum_{k=1}^{m} (a_j - \bar{a}_j)^\top S_{jk}^{-1} (a_k - \bar{a}_k)}$	Convex quadratic

Table 1: Common distance measures and their associated derived complexities. Here, a: risk variable; $\bar{a}$: baseline; S: covariance matrix.

The choice of KPI function, combined with the policy and risk variables, also influences the model complexity. Table 2 outlines three examples of standard combinations in supply chain applications. The columns present, in this order and for each example, a risk variable, a choice of KPI function, a policy, such as inventory turnover (Gaur et al., 2005; Demeter and Matyusz, 2011), fill rate (Song, 1998; Lu et al., 2003), and cost stability (i.e., variance minimization) (Choi et al., 2008; Wei and Choi, 2010), and the resulting complexity of the chosen KPI with respect to the policy π and risk variable.

The table shows that, for most examples, the KPI functions are linear in $\pi(\mathbf{a})$, which is the typical complexity result for policies commonly used in supply chain models, since policies such as order-up-to and (s, S) rules are characterized by piecewise-linear decision rules. In an order-up-to policy, inventory is replenished up to a predefined target S each period (Arrow, 1958), while a (s, S) policy places an order only when inventory falls below a threshold s, bringing it up to the maximum level S (Arrow et al., 1951). As a result, these policies can be easily formulated as mixed-integer linear programming models. Only a limited set of policies, such as the EOQ policy, introduce nonlinearities, resulting in convex quadratic KPI functions. The EOQ policy calculates the order quantity through a square-root expression that balances ordering and holding costs (Harris, 1913).

Note that the presented KPI functions conform to our definition in which low values equal bad performance and high values equal good performance. However, for other related KPI functions, such as cost variance or lead time, this rule does not hold, but an increasing form is easily obtained through simple reformulation, e.g., using cost stability instead of cost variance. Such a reformulation does not affect the complexity.

Risk variable (a)	**KPI**	**Policy**	**Complexity**
Demand	Inventory turnover: $f(\mathbf{a}, \pi(\mathbf{a})) = \frac{\sum_t a_t}{\frac{1}{T}\sum_t I_t(a)}$	Order-up-to policy: $I_{t+1}(a) = I_t(a) - a_t + \pi_t$, where $\pi_t(a) = \max\{0, S - I_t(a)\}$.	Linear in $\pi(\mathbf{a})$
Demand	Fill rate: $f(\mathbf{a}, \pi(\mathbf{a})) = \frac{\sum_t \min\{a_t, I_t(a)\}}{\sum_t a_t}$	(s, S) policy: $I_{t+1}(a) = I_t(a) - a_t + \pi_t(a)$, where $\pi_t(a) = \begin{cases} S - I_t(a), & I_t(a) \leq s, \\ 0, & I_t(a) > s. \end{cases}$	Linear in $\pi(\mathbf{a})$
Cost	Cost stability: $f(\mathbf{a}, \pi(\mathbf{a})) = \frac{1}{1 + \mathrm{Var}(\sum_t a_t Q_t(a))}$	EOQ policy: $\pi_t(a) = \sqrt{\frac{2Kd_t}{a_t}}$	Convex quadratic in $\pi(\mathbf{a})$

Table 2: Common KPI functions and their associated derived complexities. Here, I: inventory level, s: reorder point, S: order-up-to level, Q: order quantity, K: fixed ordering cost.

Overall, a natural choice for the distance measure, as summarized in Table 1, is either linear or convex quadratic. KPI functions, as shown in the first two rows of Table 2, are typically piecewise-linear. When the distance measure is linear, the resulting model can be formulated as a mixed-integer linear model, with its overall complexity mostly determined by the piecewise-linear structure that influences the number of binary variables. This ensures that the reverse stress testing model remains computationally tractable, resulting in either piecewise-linear programs or piecewise-linear programs with a quadratic objective, which can be efficiently solved by modern optimization solvers such as MOSEK, Gurobi, and CPLEX.

4.2 Criticality guarantee of reverse stress scenario

In forward stress testing, it is unclear whether any of the pre-defined input scenarios results in one of the KPI dropping below the critical threshold. If some of the predefined scenarios do, then they could far exceed thresholds and it is remains unclear what closer, more realistic scenarios are on the verge of critical thresholds. For reverse stress testing we are, under very mild assumptions, guaranteed to find a stress scenario that is precisely on the tipping point for (at least one of) the KPI metrics, i.e., we find a critical scenario. In particular, the only basic properties that should be satisfied is that: 1. in the broad set of possible realizations, such as the set of all nonnegative numbers which includes theoretically extreme scenarios, there exist a scenario that is (far) below the threshold of the KPI values; 2. the baseline scenario is not a stress scenario itself.

[C1] There exists an $\mathbf{a}$ in S for which $f_i(\mathbf{a}, \pi(\mathbf{a})) \leq q_i$ for at least one $i = 1, \ldots, m$.

[C2] The KPI functions $f_i(\mathbf{a}, \pi(\mathbf{a}))$ and the distance function $g(\mathbf{a}, \bar{\mathbf{a}})$ are continuous and closed on S for a given $\pi^*(\mathbf{a})$ and the set S itself is closed.

[C3] The baseline scenario $\bar{\mathbf{a}} \in S$ does not lie in the set F.

The first condition states that a scenario can be constructed that lies below the KPI threshold, however unlikely it seems. This is often easily checked with 'armageddon'-type scenarios such as a complete collapse or massive spike of demand. The properties stated in the second condition are more technical in nature, but hold for most used KPIs in supply chain management that are continuous as long as policies are continuous in the risk variables as well. This is the case for all KPI in Table 2 and affine or piecewise linear policies such as order-up-to policies. Notable cases where policies are not continuous are when only limited discrete batch sizes can be ordered. In all practical cases, sets S are closed, for example, the set of all real numbers or any closed interval. Finally, these properties also hold for most practical choices for g, especially when it is defined by standard distance measures such as ℓ_p-norms, which are inherently continuous. The last condition assumes that the baseline scenario does not generate a stress scenario which can often be easily tested by calculating the KPI values for the baseline value. Normally this is a formality, because one is only interested in testing policies that at least withstand the baseline situation.

The following theorem shows that a stress scenario can always be identified using the reverse stress testing optimization model, and under conditions of convexity the stress scenarios are exactly equal to one KPI value.

Theorem 1. *Under conditions (C1) and (C2), it is guaranteed that the problem in* (2) *has an optimal solution, i.e. a critical scenario closest to baseline exists and can be identified with reverse stress testing.*

Proof. Proof. Since S is a closed set and $f_i(\mathbf{a}, \pi(\mathbf{a}))$ are closed functions for all $i = 1, \ldots, m$ by condition (C2), we have that

$$F = \{\mathbf{a} \in S \mid \exists i \; f_i(\mathbf{a}, \pi(\mathbf{a})) \leq q_i, \; i = 1, \ldots, m\}$$
$$= S \cap \bigcup_{i=1}^{m} \{\mathbf{a} : f_i(\mathbf{a}, \pi(\mathbf{a})) \leq q_i\},$$

is closed as well. By condition (C1) we have that F is nonempty. Take any $\mathbf{b} \in F$ and consider the set

$$Y = \{\mathbf{a} \in F : g(\mathbf{a}, \bar{\mathbf{a}}) \leq g(\mathbf{b}, \bar{\mathbf{a}})\}.$$

Then we have that $Y = F \cap B_{g(\mathbf{b},\bar{\mathbf{a}})}$, where the latter set $B_{g(\mathbf{b},\bar{\mathbf{a}})}$ is the ball with values $\mathbf{a}$ that are within distance $g(\mathbf{b}, \bar{\mathbf{a}})$ of the origin. Note that either $\mathbf{b}$ is the infimum, or the optimal value is closer to $\bar{\mathbf{a}}$ than $\mathbf{b}$ under distance measure g. Since $B_{g(\mathbf{b},\bar{\mathbf{a}})}$ is closed and bounded, Y is also closed and bounded and we can now state

$$\inf_{\mathbf{a} \in F} g(\mathbf{a}, \bar{\mathbf{a}}) = \inf_{\mathbf{a} \in Y} g(\mathbf{a}, \bar{\mathbf{a}})$$
$$= \min_{\mathbf{a} \in Y} g(\mathbf{a}, \bar{\mathbf{a}})$$
$$= \min_{\mathbf{a} \in F} g(\mathbf{a}, \bar{\mathbf{a}}),$$

where the first equality holds since the infimum value is guaranteed to be within distance $g(\mathbf{b}, \bar{\mathbf{a}})$, so the infimum must be in the set Y. The second equality holds because g is a continuous function, so the infimum on a closed and bounded set exists and is attained according to the Weierstrass theorem. This immediately implies that the problem (2) has a solution, confirming the existence of a critical scenario. □

4.3 Robustness assessment of policy

Once a critical scenario has been identified, decision makers must assess whether the observed performance under this scenario is acceptable given their current policy. To facilitate the evaluation of this acceptability, we introduce guarantees for which scenarios are *not* critical and therefore remain within the stated boundaries of the KPI.

Let Γ^* be the optimal objective value of the reverse stress test model (2) reflecting the distance between the critical scenario and the baseline scenario. Given that no closer scenario **a** could be found that violates one of the KPI functions, this directly implies that any scenario a that is closer to the baseline scenario $\bar{a}$, the KPI values stay within the critical thresholds. From that criteria, we can say that the policy π survives any stress scenario closer within distance Γ^*.

Next to identifying the scenario **a** closest to baseline $\bar{\mathbf{a}}$ for which at least one KPI crosses their critical threshold, we can also provide guarantees for robustness of the given policy. In particular, consider the following robust optimization model, which seeks to identify an optimal policy such that KPI values remain above their critical threshold for all scenarios **a** within a distance Γ^* from the baseline $\bar{\mathbf{a}}$:

$$\min_{\pi(\cdot)} \max_{\mathbf{a}\in\mathcal{U}} \quad h\left(\pi(\mathbf{a})\right) \tag{4a}$$

$$\text{s.t.} \quad f_i(\mathbf{a}, \pi(\mathbf{a})) > q_i, \tag{4b}$$
$$\forall i = 1, \dots, m, \ \forall \mathbf{a} \in \mathcal{U}$$

where $h : \mathcal{X} \to \mathbb{R}$ is any objective function depending on the robust policy $\pi : S \to \mathcal{X}$. In robust optimization, a delicate task is the construction of the uncertainty set $\mathcal{U}$, which is given by:

$$\mathcal{U} = \{\mathbf{a} \in S : g(\mathbf{a}, \bar{\mathbf{a}}) < \Gamma\}. \tag{5}$$

Note that model (4) has an infinite number of constraints (for each $\mathbf{a} \in \mathcal{U}$). Finding such a policy is in general very hard both in theory and practice (Ben-Tal et al., 2004). However, when testing an existing supply chain, the policy π is given. In the theorem below we guarantee robustness by showing that all KPI values remain within their threshold for all scenarios in $\mathcal{U}$.

Theorem 2. *If for a given policy π, Γ is the optimal objective value of* (2)*, then π is also a feasible solution to the optimization model* (4).

Proof. Proof. For proving feasibility of π in (4), we argue by contradiction. Suppose there exist an index i and scenario $\hat{\mathbf{a}} \in \mathcal{U}(\Gamma^*)$ such that

$$f_i(\hat{\mathbf{a}}, \pi(\hat{\mathbf{a}})) \leq q_i,$$

by the definition of $\mathcal{U}(\Gamma^*)$ we also have $g(\hat{\mathbf{a}}, \bar{\mathbf{a}}) < \Gamma^*$. This contradicts the optimality of Γ^*, as no closer scenario should violate a KPI. Hence, no such $\hat{\mathbf{a}}$ exists, and the policy π satisfies

$$f_i(\hat{\mathbf{a}}, \pi(\hat{\mathbf{a}})) > q_i, \quad \forall \mathbf{a} \in \mathcal{U}(\Gamma^*), \ i = 1, \dots, m.$$

Hence, π is robustly feasible for (4). □

In particular, if we choose for g the l_2 norm and $S = \mathbb{R}^L$, then this implies that π is robustly feasible for the ellipsoidal uncertainty set $\mathcal{U} = \{a : \|\mathbf{a} - \bar{\mathbf{a}}\|_2 < \Gamma\}$ as introduced in the seminal paper on robust optimization by Ben-Tal and Nemirovski (1998).

5 Reverse stress testing applications

This section presents reverse stress testing in two types of supply chains. First, in Section 5.1, we assess a serial supply chain, where companies operate under inventory policies that determine order quantities based on past sales data. Second, in Section 5.2, we assess a service guarantee supply chain that was originally proposed by Simchi-Levi et al. (2018). This type of supply chain aims to ensure that supply shortages remain within acceptable limits. These applications illustrate how reverse stress testing works in supply chains, show that the model can be solved efficiently, and demonstrate the usefulness of reverse stress testing in uncovering supply chain vulnerabilities.

The experiments in this section are coded in Python 3.13.5 using Gurobi 12.0.3. The data and code used have been made accessible on the Github page of the first author (Gultekin, 2026).

5.1 Serial supply chain

We consider a common serial supply chain model (Lee and Whang, 1999; Lin et al., 2004; Zhang, 2005; Cachon et al., 2007; Ben-Tal et al., 2009a). This model represents a network of connected companies such as suppliers, manufacturers, warehouses, and retailers, each holding inventory, passing goods to the next, and constrained by lead times. Despite their simplicity, serial supply chains are important because they form the basis of more complex networks (Ben-Tal et al., 2009a). A key feature of this type of supply chain is the focus on optimizing the entire network rather than individual stages. A graphical overview of this structure is shown in Figure 2, highlighting the sequential flow of products through the chain.

Consider J suppliers working to fulfill the demand d_t of the end-product in $t \in \mathcal{T}^{(1)}$ time periods. Each supplier $j \in \mathcal{N} \setminus \{J\}$ orders an amount of x_{jt} from its direct upstream predecessor $j+1$ in the chain at time t based on its own inventory I_{jt} and desired inventory level, consisting of the historic sales over K periods. Note that it only bases its calculation on sales data, as we assume lost demand is not recorded. Because of lead times, it receives the goods $s_{(j+1)(t+l)}$ l periods later and only when the inventory $I_{j(t+1)}$ of its predecessor was sufficient. For this example, we take as the target inventory level the known downstream orders from the last K periods.

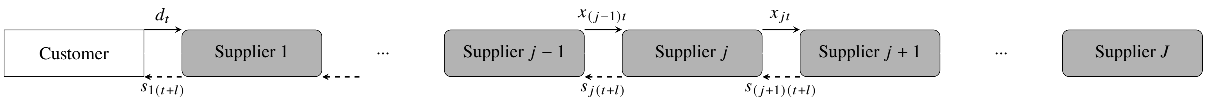


Figure 2: Serial supply chain.

5.1.1 Reverse stress testing model

The parameters and decision variables of the reverse stress testing model are provided in Table 3. Consider the uncertain demand scenario $\mathbf{d} \in \mathbb{R}_+$ to be the risk variable, and for this scenario, the policy $\pi(\mathbf{d})$ to be represented by the reorder decisions x_{jt}, supply quantities s_{jt}, and inventory levels I_{jt} for each supplier $j \in \mathcal{N}$ and time period $t \in \mathcal{T}$:

$$\pi(\mathbf{d}) := \left\{ x_{jt} \quad = \max\left\{ \textstyle\sum_{s=t-K}^{t} s_{js} - I_{jt}, 0 \right\}, \right. \tag{6}$$

where

$$s_{j(t+l)} = \begin{cases} \min\left\{ x_{(j-1)t}, I_{jt} + s_{(j+1)(t+l)} \right\} & \text{if } j \geq 2, \\ \min\left\{ d_t, I_{1t} + s_{2(t+l)} \right\} & \text{if } j = 1 \end{cases} \tag{7}$$

$$I_{jt} = \begin{cases} I_{j(t-1)} - s_{jt} + s_{(j+1)t} & \text{if } j < J, \\ I_{J(t-1)} - s_{Jt} & \text{if } j = J \end{cases}. \tag{8}$$

I_{j0} denotes the starting inventory of supplier $j \in \mathcal{N}$ and the most upstream component (e.g. raw material) is assumed to be infinite. The order quantity x_{jt} is the difference between desired inventory, based on past

Table 3: Notation for reverse stress testing in the serial supply chain model.

Sets	
$\mathcal{N}$	Set of suppliers: $\{1, ..., J\}$
$\mathcal{T}$	Set of time periods: $\{1, ..., T\}$
$\mathcal{T}^{(1)} \subseteq \mathcal{T}$	Set of feasible ordering time periods: $\{1, ..., T-l\}$
$\mathcal{T}^{(2)} \subseteq \mathcal{T}$	Set of feasible ordering periods after the initial K periods: $\{K+1, ..., T-l\}$
Parameters	
J	Number of suppliers
T	Number of time periods
l	Lead time
I_{j0}	Starting inventory of supplier j
K	Number of past periods used to determine desired inventory
q	Critical demand fulfillment level
$\bar{d}_t$	Demand in period t in the baseline scenario
Variables	
d_j	Demand in period t
x_{jt}	Order quantity of supplier j in period t from $j+1$
s_{jt}	Supply quantity sent by supplier j in period t
I_{jt}	Inventory level of supplier j in period t

demand, and the current inventory. Orders do not include any foresight on future demand. The supply $s_{j,t+l}$ is limited by the availability of goods from upstream and by either customer demand (for $j = 1$) or downstream orders (for $j \geq 2$). Finally, the inventory I_{jt} is based on the difference between outgoing and incoming flows, and no backlog is allowed.

We consider a single KPI that, for a demand scenario $d_t = (d_1, \ldots, d_T)$, calculates the fraction of orders that were not fulfilled:

$$f(\mathbf{d}, \pi(\mathbf{d})) = \frac{\sum_{t=1}^{T-l} s_{1(t+l)}}{\sum_{t=1}^{T-l} d_t}.$$

The set of demand scenarios that result in KPI values below q are given by

$$F = \left\{ \mathbf{d} \in \mathbb{R}_+^T : \frac{\sum_{t=1}^{T-l} s_{1(t+l)}}{\sum_{t=1}^{T-l} d_t} \leq q \right\}. \tag{9}$$

We seek a scenario $\mathbf{d}$ that satisfies the critical threshold q while remaining as close to the baseline demand as possible. The distance of the scenario from the baseline demand is measured as the sum of squared deviations, also known as the norm l_2:

$$g_2(\mathbf{d}, \bar{\mathbf{d}}) = \|\mathbf{d} - \bar{\mathbf{d}}\|_2. \tag{10}$$

The reverse stress test model (2) applied to the serial supply chain consisting of the policy (6), the set of stress scenarios (9), and the distance measure (10) yields a piecewise linear model, which can be equivalently reformulated as a mixed integer linear model, see Appendix B.

5.1.2 Numerical experiment

We consider the baseline demand scenarios $\bar{\mathbf{d}}$ over $T - l = 17$ weeks, where $T = 18$ and $l = 1$, as depicted in Figure 3 which are derived using

$$\bar{d}_t = 1000 \left(1 + \frac{1}{2} \sin\left(\frac{\pi(t-1)}{4}\right)\right), \quad t = 1, \ldots, 17,$$

which reflects a seasonal demand pattern, with its peak in the beginning. To avoid an unrealistic drop in inventory in the first K periods, where historical sales data are unavailable, we assume a constant number

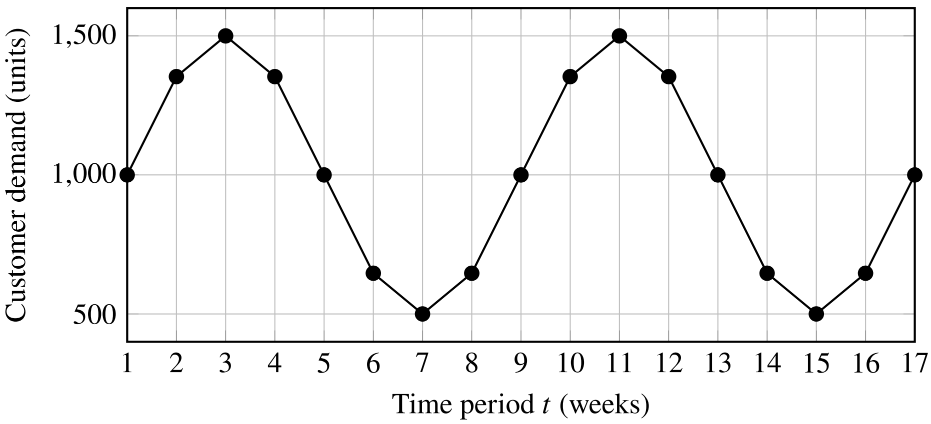


Figure 3: Baseline demand scenario for serial supply chain.

of orders, equal to $x_{jt} = I_{0,j}/K$ coming in the first $t = 1, \ldots, K$ periods for each supplier $j \in \mathcal{N} \setminus \{J\}$. The chosen values of these parameters are given in Table 4.

Parameter	Value
J	5
T	18
l	1
I_{j0}	1000
K	2
q	0.5

Table 4: Parameter values for the numerical example ($I_{j0} = 1000, \quad \forall j \in \mathcal{N} \setminus \{J\}$).

5.1.3 Results

In the numerical experiments, we aimed to demonstrate the reverse stress testing model to identify the critical demand scenario that causes the demand fulfillment KPI to fall below the threshold. All runs were solved to optimality within ten minutes, demonstrating that the model is computationally efficient.

We first measured the performance for the baseline demand scenario. For the selected policy $\pi(\mathbf{d})$, the baseline scenario achieved a demand fulfillment rate of 92.9%. The reason for missing out on full demand fulfillment in the baseline scenario was due to the initial constant ordering in the first K periods and missing a small portion of demand in those first periods. The observed performance demonstrated that the policy performed satisfactorily under normal conditions.

Second, we identified the critical scenario and analyzed how it deviated from the baseline pattern. This scenario reduced the KPI below the critical threshold of 50%. Figure 4 shows both the baseline and the critical scenarios. The baseline scenario exhibited a seasonal pattern with alternating peaks and troughs between 500 and 1,500 units. In contrast, the critical scenario identified the minimal deviations from this baseline that were sufficient to reduce the fulfillment rate below the critical threshold. It started with unusually low demand in the early periods, followed by a steep ramp up in demand in periods 9-13. After this surge, demand declined steeply before partially recovering in the last two periods. The most important observation was that demand peaked in the later periods, particularly weeks 9–13, which were critical, as they could exceed available inventory and create shortages. This phenomena of steep ramps and their impact on supply chains are common in long or complex supply chain such as the semiconductor supply chain.

Figure 5 shows the order quantities of each supplier over time under the critical scenario. In the first two periods, orders were constant for all suppliers due to the initial ordering constraint ($x_{jt} = I_{0,j}/K$). The *bullwhip effect* is clearly visible in the ordering quantities: small variations in customer demand are amplified as they propagate upstream, leading to larger fluctuations in supplier orders. The amplification

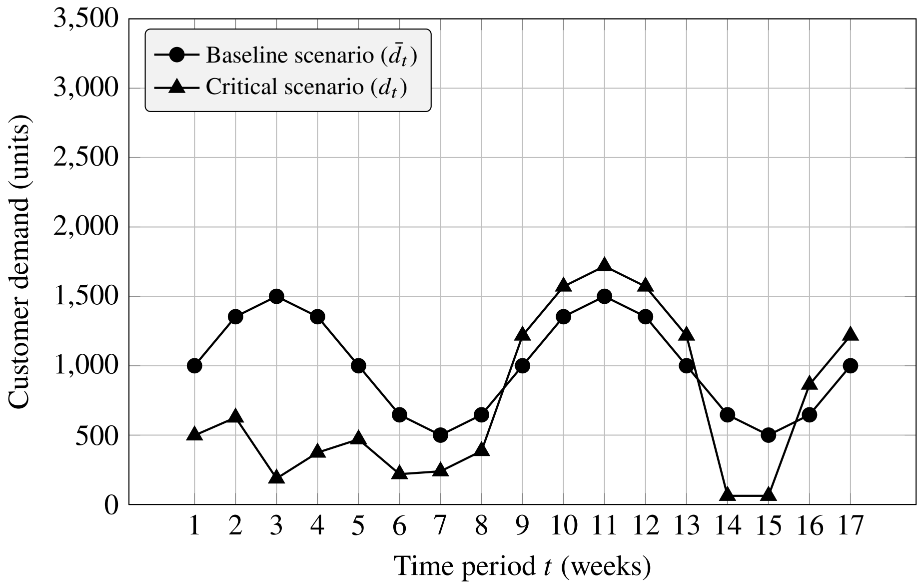


Figure 4: Demand deviations from the baseline over time.

was uneven across suppliers. Supplier 1 reacted early with the largest initial increase, while supplier 4 reacted later but experienced the steepest rise in orders. This highlights that different suppliers may have experienced peak demand at different times, increasing upstream risks. The pattern also indicates which suppliers were most vulnerable. Supplier 1 was highly sensitive in the early periods, and supplier 4 became critical later.

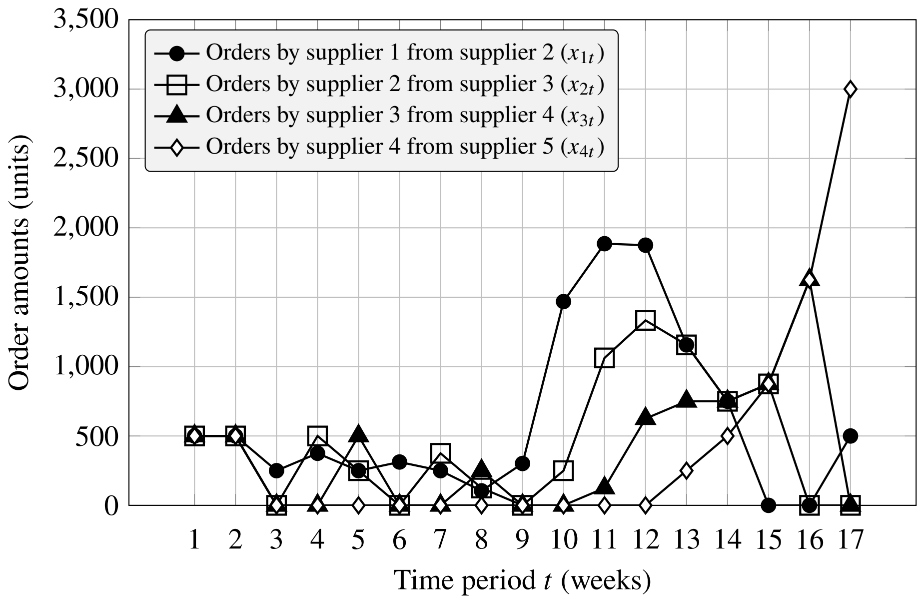


Figure 5: Order amounts over time.

5.1.4 Multiple stress scenarios

So far, our analyses have relied on the l_2 norm as the primary distance measure. To explore alternative stress scenarios, we used different distance measures to compare against the baseline. Figure 6 illustrates the impact of different distance measures on demand. The baseline demand is shown alongside the l_1, l_2, and l_∞ norm critical scenarios. While the l_1 and l_2 norms produced somewhat similar overall demand patterns, the l_1 norm resulted in more localized, sharper changes in specific periods (e.g., period 13). The l_∞ norm spread deviations evenly but forced large, unrealistic changes. By contrast, the l_2 norm spread deviations smoothly across multiple periods, limiting extreme spikes and creating a more balanced and realistic demand scenario. This makes the l_2 norm particularly suitable for reverse stress testing, as it captures systemic vulnerabilities in a practically meaningful way.

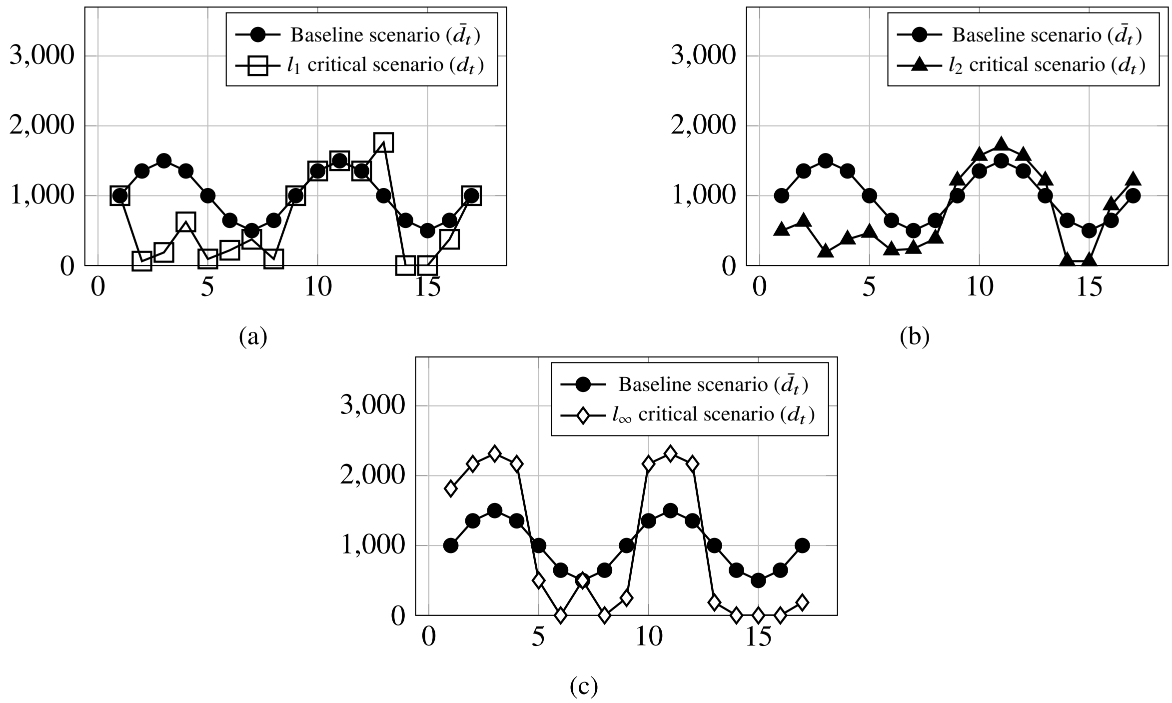


Figure 6: Sensitivity of the critical scenario to the distance measures: (a) l_1 norm, (b) l_2 norm, and (c) l_∞ norm. The y-axes show customer demand (units) and the x-axes show time periods (weeks).

The second approach to generating alternative stress scenarios involves randomly searching the neighborhood of the initial optimal solution, as outlined in the model (3). Let Γ denote the optimal objective value obtained from the initial reverse stress testing model in (10). We solve the model iteratively by generating random objective direction vectors c, where each element is uniformly drawn from the interval $[-1, 1]$. To obtain a meaningful variety of stress scenarios, we search for scenarios that could be twice as far in distance from the baseline as the critical scenario. This is equivalent to setting $\epsilon = 1$ in (3). While we initially tested smaller values of ϵ, we found that they restricted the search space too severely and failed to produce sufficiently diverse demand patterns. The results for smaller ϵ values are provided in Appendix F.

Solving this modified model is computationally demanding, as moving the quadratic distance function $g(.)$ from the objective into the constraints increases the problem's complexity. To handle this complexity, we applied a 10-minute time limit for each iterative run. Furthermore, to assist the solver in finding a feasible solution quickly, we warm-started the model by providing the l_∞ critical scenario (previously illustrated in Figure 6) as an initial feasible solution. Consequently, as depicted in Figure 7, the solver successfully generated a diverse set of distinct demand patterns within the time limits, providing decision makers with a broader perspective on supply chain vulnerabilities.

5.2 Service guarantee supply chain

We next explore a service guarantee supply chain model based on Simchi-Levi et al. (2018) where they consider a supply chain with M plants, where each plant is denoted by $\mathcal{S}_i$ for $i = 1, \ldots, M$, and N products, where each product is denoted by $\mathcal{T}_j$ for $j = 1, \ldots, N$. A plant may be dedicated to a single product or equipped to produce multiple products, depending on the flexibility design that outlines its

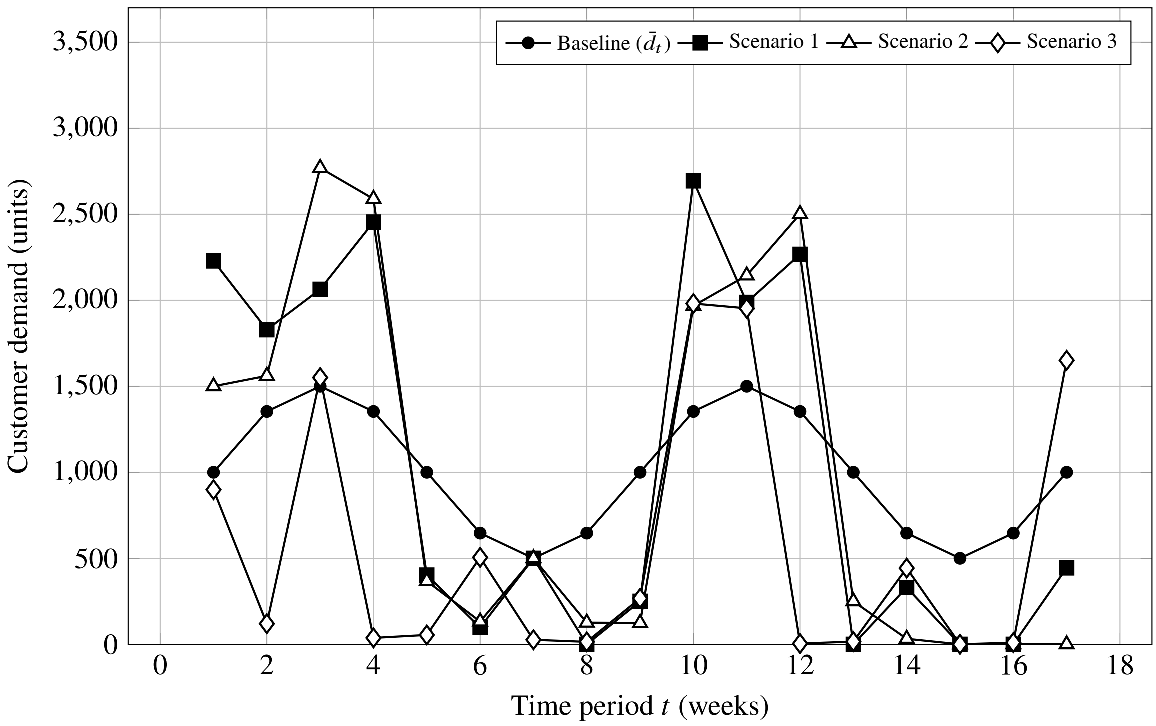


Figure 7: Baseline and multiple stress scenarios obtained from different random iterations. ($\epsilon = 1$)

production capabilities. The service guarantee (SG) model guarantees that the total demand loss (l_j) remains within the predetermined shortage allowance, δ_1, while aiming to minimize inventory costs. In addition, it determines the levels of the inventory of products (s_j) and the production schedules of plants (x_{ij}).

5.2.1 Reverse stress testing model

The sets, parameters, and decision variables of the reverse stress testing formulation for the service guarantee (RST-SG) model are provided in Table 5.

Table 5: Notation for reverse stress testing in the service guarantee supply chain model.

Sets	
$\mathcal{S}_i$	Plants ($i \in \{1, \dots, M\}$)
$\mathcal{T}_j$	Products ($j \in \{1, \dots, N\}$)
Parameters	
$\bar{d}_j$	Demand of product j in baseline scenario
$\bar{c}_i$	Capacity of plant i in baseline scenario
$\bar{x}_{ij}$	Quantity of product j produced by plant i in baseline scenario
$\bar{s}_j$	Inventory level of product j in baseline scenario
δ_1	Shortage allowance in the service guarantee model
δ_2	Shortage allowance in the reverse stress testing model (critical threshold)
Variables	
d_j	Demand of product j
c_i	Capacity of plant i
x_{ij}	Quantity of product j produced by plant i
l_j^d	Lost demand of product j
l_i^c	Capacity shortage of plant i

In this model, we consider the uncertain capacity and demand $(\mathbf{c}, \mathbf{d}) \in \mathbb{R}_+^M \times \mathbb{R}_+^N$ to be the risk variables, and the policy $\pi(\mathbf{d})$ is determined by the allocation decisions x_{ij}:

$$\pi(\mathbf{d}) = \left\{ x_{ij} = \frac{d_j}{\bar{d}_j}\bar{x}_{ij} \quad i = 1, \dots, M, j = 1, \dots, N \right. \tag{11}$$

The policy idea ($x_{ij} = (d_j/\bar{d}_j)\bar{x}_{ij}$) is based on the proportional demand model from the robust facility location model in Baron et al. (2011). We substitute this value for $x_{ij}(\mathbf{c}, \mathbf{d})$ in the model.

We consider a KPI that evaluates the total demand and capacity losses for a scenario $(\mathbf{c}, \mathbf{d})$ as follows:

$$f((\mathbf{c}, \mathbf{d}), \pi(\mathbf{c}, \mathbf{d})) = \sum_{j=1}^{N} l_j^d(\mathbf{c}, \mathbf{d}) + \sum_{i=1}^{M} l_i^c(\mathbf{c}, \mathbf{d}),$$

where

$$l_j^d = max\{d_j - \bar{s}_j - \sum_{i=1}^{M} \frac{d_j}{\bar{d}_j}\bar{x}_{ij}, 0\} \qquad j = 1, \dots, N \tag{12}$$

$$l_i^c = max\{\sum_{j=1}^{N} \frac{d_j}{\bar{d}_j}\bar{x}_{ij} - c_i, 0\} \qquad i = 1, \dots, M. \tag{13}$$

The lost demand variable l_j^d captures the unmet demand of product j resulting from limited supply, whereas the capacity loss variable l_i^c reflects the shortage caused by insufficient capacity at supplier i.

The set of scenarios that lead to KPI values exceeding the critical threshold δ_2 is defined as:

$$F = \Big\{ (\mathbf{c}, \mathbf{d}) \in \mathbb{R}_+^M \times \mathbb{R}_+^N : \sum_{j=1}^{N} l_j^d(\mathbf{c}, \mathbf{d}) + \sum_{i=1}^{M} l_i^c(\mathbf{c}, \mathbf{d}) \geq \delta_2 \sum_{j=1}^{N} \bar{d}_j \Big\} \tag{14}$$

The direction of the inequality differs from that in (1), since the model in Simchi-Levi et al. (2018) specifies the KPI as total loss. Therefore, (14) is formulated using the "≥" direction to ensure consistency with this definition.

The distance between the scenario and the baseline parameters is calculated as the l_2 norm:

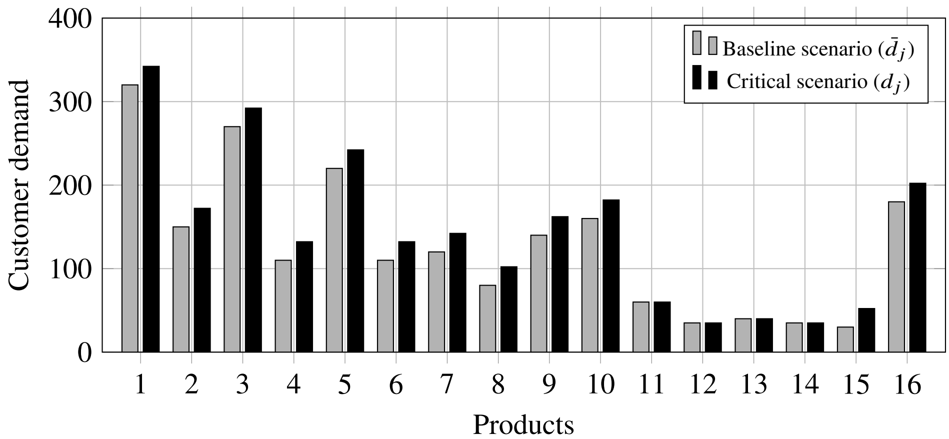
$$g_2((\mathbf{c}, \mathbf{d}), (\bar{\mathbf{c}}, \bar{\mathbf{d}})) = \|\mathbf{c} - \bar{\mathbf{c}}\|_2 + \|\mathbf{d} - \bar{\mathbf{d}}\|_2. \tag{15}$$

For this application, the reverse stress test model (2), appended with the specific policy (11), set of stress scenarios (14), and distance measure (15) can be equivalently formulated as a mixed-integer linear model, as detailed in the Appendix C.

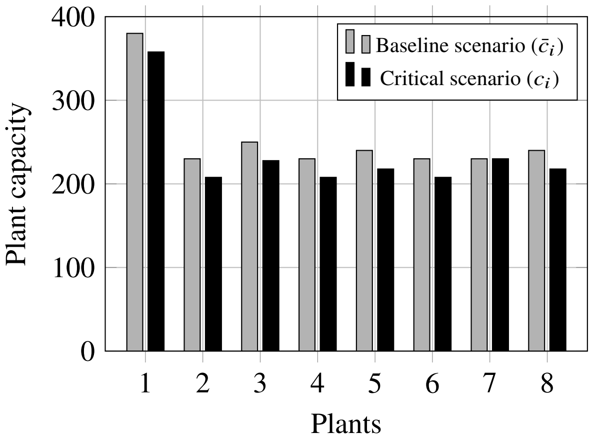
5.2.2 Numerical experiment

We implement our framework on a case study involving automobile supply chain presented by Simchi-Levi et al. (2018). The dataset consists of 16 different vehicle types and 8 assembly plants. We adopt the same plant capacities ($\bar{c}_i$) and average product demands ($\bar{d}_j$) as outlined in their study. The case study includes three flexibility designs: 2-chain, 3-chain, and full flexibility designs. We solve the SG model inspired by Simchi-Levi et al. (2018) to obtain the baseline scenario parameters, which then serve as inputs for RST-SG model. To obtain the baseline scenario, we employ the SG model with the parameter δ_1 set to zero, ensuring no demand loss is allowed. For the SG model and further details regarding the derivation of the baselines, see Appendix D. For the RST-SG model we use a critical threshold δ_2 of 20%. Additionally, we assume a uniform unit holding cost for all products, as in the original paper.

5.2.3 Results



(a)



(b)

Figure 8: Deviations from the baseline for the 2-chain design: (a) product demand, (b) plant capacity.

In the numerical experiments, we identified the critical scenario in which total shortages exceed the threshold. The model produced results within seconds, confirming its computational efficiency.

Unlike Simchi-Levi et al. (2018), who employed robust optimization to identify optimal decisions within a predefined uncertainty set, our model differs conceptually. We first formalized a decision policy and then tested the supply chain's resilience without relying on an explicit uncertainty set. Specifically, we adopted a simple proportional allocation rule as the policy $\pi(\mathbf{d})$. Despite its simplicity, the resulting critical scenario was equivalent to the optimal response under this rule. To verify this, we also resolved the SG model for the critical scenario and obtained the exact allocation, confirming that the proportional policy was the best response in this case.

We then identified the critical scenario, highlighting the deviations from the baseline across products and plants for the 2-chain design, as shown in Figure 8. The corresponding and comparable results for the 3-chain design are provided in Appendix E. On the product side, shown in Figure 9a, high-demand products increased by about 10%, while low-demand products 11,12, 13, and 14 remained unchanged. On the plant side, shown in Figure 9b, most critical plants experienced reductions of approximately 10%, whereas plant 7 remained unaffected. Plant 7 and its connected low-demand products were not stressed because their capacities were high and substantial reductions would be needed to exceed the shortage threshold. The scenario therefore focuses on high-demand products and critical plants where changes have the largest system-wide impact. Overall, the critical scenario amplifies pressures on the most significant products and plants, revealing the elements whose disruption has the greatest impact on supply chain performance.

6 Conclusion

This research presents reverse stress testing for supply chains that works backward to find the exact conditions that cause a supply chain to fail. Compared to forward stress testing, one does not need to rely on scenarios that are generated beforehand and estimated probabilities. In practice, such scenarios and probabilities are often hard to obtain, making forward stress testing susceptible to significant blind spots. Our method is therefore complementary to forward stress testing and other methods, including robust optimization, such as the model developed by Simchi-Levi et al. (2018), or stochastic optimization. While these methods intend to construct resilient solutions, reverse stress testing specifically evaluates those solutions by identifying the precise conditions under which they fail to meet performance requirements, potentially even beyond or before the thresholds they were intended or believed to withstand. By identifying the minimal shocks that trigger failure, we provide a formal robustness guarantee that a given solution remains feasible for all scenarios within the identified distance from the baseline. These results serve as a foundation for managerial evaluation and decision-making.

Once a stress scenario is identified, a key part of the managerial evaluation is assessing its realism to determine if the vulnerability requires mitigation or is unlikely enough to be deemed acceptable without further action. Decision makers must analyze whether the scenario is plausible by considering factors such as industry conditions, historical data, and expert judgment. This can inspire targeted adjustments in strategic planning, such as adapting procurement strategies, modifying production processes, diversifying suppliers, or developing contingency plans. While detailing these specific adjustments is beyond our scope, this study empowers managers to build long term resilience.

References

Al-Khatib, M., Haji, M., Haouari, M., and Kharbeche, M. (2024). Building resilience in the infant formula milk supply chain. *Food Control*, 165:110641.

Albanese, C., Crépey, S., and Iabichino, S. (2023). Quantitative reverse stress testing, bottom up. *Quantitative Finance*, 23(5):863–875.

Allam, Z., Bibri, S. E., and Sharpe, S. A. (2022). The rising impacts of the covid-19 pandemic and the russia–ukraine war: energy transition, climate justice, global inequality, and supply chain disruption. *Resources*, 11(11):99.

Arrow, K. J. (1958). Historical background. In Arrow, K. J., Karlin, S., and Scarf, H., editors, *Studies in the Mathematical Theory of Inventory and Production*. Stanford University Press, Stanford, CA.

Arrow, K. J., Harris, T., and Marschak, J. (1951). Optimal inventory policy. *Econometrica: Journal of the Econometric Society*, pages 250–272.

Baron, O., Milner, J., and Naseraldin, H. (2011). Facility location: A robust optimization approach. *Production and Operations Management*, 20(5):772–785.

Basel Committee on Banking Supervision (2009). Principles for sound stress testing practices and supervision. Report, Bank for International Settlements. Accessed: 10 November 2025.

Ben-Tal, A., Boaz, G., and Shimrit, S. (2009a). Robust multi-echelon multi-period inventory control. *European Journal of Operational Research*, 199(3):922–935.

Ben-Tal, A., Den Hertog, D., and Vial, J.-P. (2015). Deriving robust counterparts of nonlinear uncertain inequalities. *Mathematical programming*, 149(1):265–299.

Ben-Tal, A., Goryashko, A., Guslitzer, E., and Nemirovski, A. (2004). Adjustable robust solutions of uncertain linear programs. *Mathematical programming*, 99(2):351–376.

Ben-Tal, A. and Nemirovski, A. (1998). Robust convex optimization. *Mathematics of operations research*, 23(4):769–805.

Ben-Tal, A. and Nemirovski, A. (2002). Robust optimization–methodology and applications. *Mathematical programming*, 92(3):453–480.

Ben-Tal, A., Nemirovski, A., and El Ghaoui, L. (2009b). *Robust optimization*. Princeton university press.

Berntsen, P. B. and Trutnevyte, E. (2017). Ensuring diversity of national energy scenarios: Bottom-up energy system model with modeling to generate alternatives. *Energy*, 126:886–898.

Bertsimas, D. and den Hertog, D. (2022). *Robust and Adaptive Optimization*. Dynamic Ideas LLC, Belmont, MA.

Bertsimas, D. and Sim, M. (2004). The price of robustness. *Operations research*, 52(1):35–53.

Breuer, T. and Csiszár, I. (2013). Systematic stress tests with entropic plausibility constraints. *Journal of Banking & Finance*, 37(5):1552–1559.

Brintrup, A., Pak, J., Ratiney, D., Pearce, T., Wichmann, P., Woodall, P., and McFarlane, D. (2020). Supply chain data analytics for predicting supplier disruptions: a case study in complex asset manufacturing. *International Journal of Production Research*, 58(11):3330–3341.

Bruetzel, O., Hörger, M., Foran, A., Benfer, M., Nassehi, A., and Lanza, G. (2025). Robust planning of production networks at an automotive supplier. *International Journal of Production Research*, 63(9):3365–3383.

Cachon, G. P., Randall, T., and Schmidt, G. M. (2007). In search of the bullwhip effect. *Manufacturing & Service Operations Management*, 9(4):457–479.

Chan, T. C. and Kaw, N. (2020). Inverse optimization for the recovery of constraint parameters. *European Journal of Operational Research*, 282(2):415–427.

Chan, T. C., Mahmood, R., and Zhu, I. Y. (2025). Inverse optimization: Theory and applications. *Operations Research*, 73(2):1046–1074.

Choi, T.-M., Li, D., and Yan, H. (2008). Mean–variance analysis of a single supplier and retailer supply chain under a returns policy. *European Journal of Operational Research*, 184(1):356–376.

Chopra, S. and Sodhi, M. S. (2014). Reducing the risk of supply chain disruptions. *MIT Sloan Management Review*, 55(3):73–80.

Committee of European Banking Supervisors (2009). Guidelines on stress testing (cp32). Report, European Banking Authority. Accessed: 10 November 2025.

Demeter, K. and Matyusz, Z. (2011). The impact of lean practices on inventory turnover. *International journal of production economics*, 133(1):154–163.

Diem, C., Schueller, W., Gerschberger, M., Stangl, J., Conrady, B., Gerschberger, M., and Thurner, S. (2025). Supply network stress-testing of food security on the establishment-level. *International Journal of Production Research*, 63(9):3259–3283.

Dolgui, A., Ivanov, D., and Rozhkov, M. (2020). Does the ripple effect influence the bullwhip effect? an integrated analysis of structural and operational dynamics in the supply chain. *International Journal of Production Research*, 58(5):1285–1301.

Dolgui, A., Ivanov, D., and Simchi-Levi, D. (2025). Stress tests for supply chains: towards resilience and viability.

Financial Services Authority (2008). Stress and scenario testing: cp08/24. Consultation paper, Financial Services Authority. Accessed: 10 November 2025.

Gao, S. Y., Simchi-Levi, D., Teo, C.-P., and Yan, Z. (2019). Disruption risk mitigation in supply chains: The risk exposure index revisited. *Operations Research*, 67(3):831–852.

Gaur, V., Fisher, M. L., and Raman, A. (2005). An econometric analysis of inventory turnover performance in retail services. *Management science*, 51(2):181–194.

Gorissen, B. L., den Hertog, D., and Reusken, M. (2025). Hidden convexity in a class of optimization problems with bilinear terms. *Operations Research*.

Grigat, D. and Caccioli, F. (2017). Reverse stress testing interbank networks. *Scientific reports*, 7(1):15616.

Grundke, P. (2011). Reverse stress tests with bottom-up approaches. *The Journal of Risk Model Validation*, 5(1):71–90.

Grundke, P. and Pliszka, K. (2018). A macroeconomic reverse stress test. *Review of Quantitative Finance and Accounting*, 50:1093–1130.

Guan, D., Wang, D., Hallegatte, S., Davis, S. J., Huo, J., Li, S., Bai, Y., Lei, T., Xue, Q., Coffman, D., et al. (2020). Global supply-chain effects of covid-19 control measures. *Nature human behaviour*, 4(6):577–587.

Gultekin, B. (2026). Reverse stress testing for supply chains [source code]. `https://github.com/bgultekiin/Reverse-stress-testing-for-supply-chains`.

Gunasekaran, A., Subramanian, N., and Rahman, S. (2015). Supply chain resilience: role of complexities and strategies. *International Journal of Production Research*, 53(22):6809–6819.

Haraguchi, M. and Lall, U. (2015). Flood risks and impacts: A case study of thailand's floods in 2011 and research questions for supply chain decision making. *International Journal of Disaster Risk Reduction*, 14:256–272.

Harris, F. W. (1913). What quantity to make at once. In *The Library of Factory Management: Operation and Costs*, volume 5 of *The Factory Management Series*, pages 47–52. A. W. Shaw Co., Chicago.

Hollis, S. and Ekengren, M. (2025). The reality of contingency: Implications for crisis management. *Risk, Hazards & Crisis in Public Policy*, 16(3):e70020.

Hosseini, S. and Ivanov, D. (2022). A new resilience measure for supply networks with the ripple effect considerations: A bayesian network approach. *Annals of Operations Research*, 319(1):581–607.

Hosseini, S., Ivanov, D., and Dolgui, A. (2019). Review of quantitative methods for supply chain resilience analysis. *Transportation research part E: logistics and transportation review*, 125:285–307.

Hosseini, S., Ivanov, D., and Dolgui, A. (2020). Ripple effect modelling of supplier disruption: integrated markov chain and dynamic bayesian network approach. *International Journal of Production Research*, 58(11):3284–3303.

Ivanov, D. (2017). Simulation-based ripple effect modelling in the supply chain. *International Journal of Production Research*, 55(7):2083–2101.

Ivanov, D. (2020). Predicting the impacts of epidemic outbreaks on global supply chains: A simulation-based analysis on the coronavirus outbreak (covid-19/sars-cov-2) case. *Transportation Research Part E: Logistics and Transportation Review*, 136:101922.

Ivanov, D. (2024). Exiting the covid-19 pandemic: After-shock risks and avoidance of disruption tails in supply chains. *Annals of Operations Research*, 335(3):1627–1644.

Ivanov, D. (2025). Literature review on the stress tests developed and applied in non-food supply chains. Report, Publications Office of the European Union, Luxembourg. Accessed: 10 November 2025.

Khakifirooz, M., Fathi, M., Dolgui, A., and Pardalos, P. M. (2025). Assessing resiliency in scale-free supply chain networks: a stress testing approach based on entropy measurements and value-at-risk analysis. *International Journal of Production Research*, 63(9):3331–3364.

Kinra, A., Ivanov, D., Das, A., and Dolgui, A. (2020). Ripple effect quantification by supplier risk exposure assessment. *International Journal of Production Research*, 58(18):5559–5578.

Kopeliovich, Y., Novosyolov, A., Satchkov, D., and Schachter, B. (2015). Robust risk estimation and hedging: A reverse stress testing approach. *The Journal of Derivatives*, 22(4):10–25.

Korikov, A. and Beck, J. C. (2021). Counterfactual explanations via inverse constraint programming. In *27th International Conference on Principles and Practice of Constraint Programming (CP 2021)*, pages 35–1. Schloss Dagstuhl–Leibniz-Zentrum für Informatik.

Korikov, A. and Beck, J. C. (2023). Objective-based counterfactual explanations for linear discrete optimization. In *International Conference on Integration of Constraint Programming, Artificial Intelligence, and Operations Research*, pages 18–34. Springer.

Korikov, A., Shleyfman, A., and Beck, C. (2021). Counterfactual explanations for optimization-based decisions in the context of the gdpr. In *ICAPS 2021 workshop on explainable AI planning*.

Kurtz, J., Birbil, Ş. İ., and den Hertog, D. (2025). Counterfactual explanations for linear optimization. *European Journal of Operational Research*.

Lee, H. and Whang, S. (1999). Decentralized multi-echelon supply chains: Incentives and information. *Management science*, 45(5):633–640.

Li, Y., Chen, K., Collignon, S., and Ivanov, D. (2021). Ripple effect in the supply chain network: Forward and backward disruption propagation, network health and firm vulnerability. *European Journal of Operational Research*, 291(3):1117–1131.

Li, Y. and Zobel, C. W. (2020). Exploring supply chain network resilience in the presence of the ripple effect. *International journal of production economics*, 228:107693.

Lin, P.-H., Wong, D. S.-H., Jang, S.-S., Shieh, S.-S., and Chu, J.-Z. (2004). Controller design and reduction of bullwhip for a model supply chain system using z-transform analysis. *Journal of process control*, 14(5):487–499.

Liu, C., Tang, L., Lin, D., and Guo, J. (2023). Testing to extreme: An application of reverse stress testing engineering on mortgages of commercial banks in china. *International Journal of Finance & Economics*, 28(1):187–192.

Liu, M., Ding, Y., Chu, F., Dolgui, A., and Zheng, F. (2024). Robust actions for improving supply chain resilience and viability. *Omega*, 123:102972.

Liu, M., Liu, Z., Chu, F., Zheng, F., and Chu, C. (2021). A new robust dynamic bayesian network approach for disruption risk assessment under the supply chain ripple effect. *International Journal of Production Research*, 59(1):265–285.

Lu, Y., Song, J.-S., and Yao, D. D. (2003). Order fill rate, leadtime variability, and advance demand information in an assemble-to-order system. *Operations Research*, 51(2):292–308.

Mitic, P. (2021). Operational risk reverse stress testing: Optimal solutions. *Mathematical and Computational Applications*, 26(2):38.

Montesi, G., Papiro, G., Fazzini, M., and Ronga, A. (2020). Stochastic Optimization System for Bank Reverse Stress Testing. *Journal of Risk and Financial Management*, 13(8):174.

Pettit, T. J., Fiksel, J., and Croxton, K. L. (2010). Ensuring supply chain resilience: development of a conceptual framework. *Journal of business logistics*, 31(1):1–21.

Reuters (2026). China warns global chip shortages; Nexperia dispute escalates again. Reuters. Accessed: 31 March 2026.

Sawik, T. (2022). Stochastic optimization of supply chain resilience under ripple effect: A covid-19 pandemic related study. *Omega*, 109:102596.

Simchi-Levi, D., Schmidt, W., Wei, Y., Zhang, P. Y., Combs, K., Ge, Y., Gusikhin, O., Sanders, M., and Zhang, D. (2015). Identifying risks and mitigating disruptions in the automotive supply chain. *Interfaces*, 45(5):375–390.

Simchi-Levi, D. and Simchi-Levi, E. (2020). We need a stress test for critical supply chains. *Harvard Business Review*, 28:2020.

Simchi-Levi, D., Wang, H., and Wei, Y. (2018). Increasing supply chain robustness through process flexibility and inventory. *Production and Operations Management*, 27(8):1476–1491.

Song, J.-S. (1998). On the order fill rate in a multi-item, base-stock inventory system. *Operations research*, 46(6):831–845.

Wei, Y. and Choi, T.-M. (2010). Mean–variance analysis of supply chains under wholesale pricing and profit sharing schemes. *European Journal of Operational Research*, 204(2):255–262.

Weller-Fahy, D. J., Borghetti, B. J., and Sodemann, A. A. (2014). A survey of distance and similarity measures used within network intrusion anomaly detection. *IEEE Communications Surveys & Tutorials*, 17(1):70–91.

Wicaksana, A., Ho, W., Talluri, S., and Dolgui, A. (2022). A decade of progress in supply chain risk management: risk typology, emerging topics, and research collaborators. *International Journal of Production Research*, 60(24):7155–7177.

Zhang, X. (2005). Delayed demand information and dampened bullwhip effect. *Operations Research Letters*, 33(3):289–294.

A The reverse stress testing framework

This appendix provides a detailed description of the six-step reverse stress testing framework introduced in Section 3 and visualized in Figure 1, outlining the steps involved from selecting risk variables to identifying the critical scenario.

A.1 Select risk variables

The first step in reverse stress testing is to select the risk variables that represent sources of uncertainty in the supply chain that may deviate from their expected values under adverse conditions, such as natural disasters, political instability, or economic shocks. The risk variables may vary depending on the context and scope of the supply chain, and the interest of the decision maker. Typical risk variables include demand, supply, lead time, and costs, and are characterized with some uncertainty.

A.2 Define baselines

After the risk variables are identified, the next step defines their baseline, which represents the supply chain's normal or expected operating conditions. These baseline values serve as reference points for measuring how far a scenario deviates under stress. A baseline can be based on historical averages, forecasts, or planned values, depending on the availability and quality of data, and the selected type of risk variable. For example, the baseline for demand may correspond to the forecasted demand for the upcoming period, while lead time could reflect the average duration observed under stable conditions.

A.3 Formalize policy

In this step, a policy is defined. In supply chains, a policy is a structured set of decision rules that govern operations, such as inventory replenishment strategies, supplier selection criteria, production scheduling, and stock or allocation rules. A policy depends on the risk variables. For instance, if demand is selected as a risk variable and a retailer follows a (s, S) inventory policy, this policy responds to demand fluctuations by restocking products only when inventory drops below a certain threshold s and orders up to a predetermined level S. Hence, policies determine how supply chains react to different external conditions and constraints, including predefined procedures and adaptive measures in times of crisis. These measures may involve temporary changes, such as sourcing from alternative suppliers, modifying products, reallocating inventory, or adjusting transportation strategies.

A.4 Formalize KPIs

In this step of the framework, the KPIs are selected that are most important for the decision makers managing the supply chain. These KPIs serve as benchmarks for assessing the operational and financial health of the supply chain. Commonly used KPIs in supply chain risk management include service level, lead time, and cost efficiency, each reflecting critical aspects of supply chain performance such as demand fulfillment, replenishment speed, and the financial impact of disruptions. In addition, time-based performance metrics are also important. Two key examples are Time to Recover (TTR) and Time to Survive (TTS). TTR is the time it takes for a disrupted facility, such as a supplier or warehouse, to return to normal operations. TTS is the longest time the supply chain can continue meeting demand using available inventory or backup options after a disruption (Simchi-Levi and Simchi-Levi, 2020). These indicators are KPIs that provide a quantifiable means to evaluate supply chain performance.

A.5 Determine critical thresholds

Once the KPIs have been established, the next step is to identify their critical thresholds—points at which disruptions transition from acceptable inconveniences to severe operational risks. They can be defined as cutoff values for KPIs beyond which the supply chain experiences systemic failure or significant financial losses. For instance, if demand backlog grows to a level where fulfilling it would exceed acceptable customer lead times, or if lost sales reach a volume that surpasses the company's financial buffers and threatens its commercial viability, the supply chain is considered to have crossed a critical threshold.

A.6 Identify the critical scenario

In this final step, the risk variables, baselines, policy, KPIs, and critical thresholds defined in the previous steps are combined to identify the critical scenario. The critical scenario marks the first failure point, i.e., the point closest to normal operations where the supply chain no longer meets its performance requirements. To determine this scenario, we introduce an mathematical optimization model (Section 3) that computes the minimal deviation in the risk variables from their baselines that results in KPI values breaching their thresholds.

B Mixed integer linear formulation of serial supply chain

The reverse stress test model (2) applied to the serial supply chain in Section 5.1 with the policy (6) , the distance measure (10), and the set of stress scenarios (9) is:

$$\min \quad \sum_{t=1}^{T-l}(d_t - \bar{d}_t)^2 \tag{16}$$

$$\text{s.t.} \quad \frac{\sum_{t=1}^{T-l} s_{1(t+l)}}{\sum_{t=1}^{T-l} d_t} \leq q, \tag{17}$$

$$x_{jt} = \max\{\textstyle\sum_{s=t-K}^{t} s_{js} - I_{jt}, 0\}, \quad j < J, t \in \mathcal{T}^{(2)}, \tag{18}$$

$$x_{jt} = I_{j0}/K, \quad j < J, t = 1, \ldots, l, \tag{19}$$

$$s_{1(t+l)} = \min\{d_t, I_{1t} + s_{2(t+l)}\}, \quad t \in \mathcal{T}^{(2)}, \tag{20}$$

$$s_{j(t+l)} = \min\{x_{(j-1)t}, I_{jt} + s_{(j+1)(t+l)}\}, \; j > 1, t \in \mathcal{T}^{(1)}, \tag{21}$$

$$I_{jt} = I_{j(t-1)} - s_{jt} + s_{(j+1)t}, \quad j < J, t \in \mathcal{T}, \tag{22}$$

$$I_{Jt} = I_{J(t-1)} - s_{Jt}, \quad t \in \mathcal{T}. \tag{23}$$

This model has some nonlinear constraints and some piecewise linear constraints. Below we use standard tricks from linear programming to reformulate these constraints into equivalent (integer) linear programming constraints.

The constraint (17) can be directly linearized using the following equivalence:

$$\sum_{t=1}^{T-l} s_{1(t+l)} \leq q \sum_{t=1}^{T-l} d_t,$$

as $d_t \geq 0$ for all $t \in \mathcal{T}^{(1)}$. For the piecewise linear maximization describing the order quantity we use so-called Big-M constraints. To do so, we introduce additional binary variables $v_{jt} \in \{0, 1\}$, $j < J$, $t \in \mathcal{T}^{(1)}$ and replace the constraints (18) by the following set of constraints

$$x_{jt} \geq 0, \quad j < J, \; t \in \mathcal{T}^{(2)},$$

$$x_{jt} \geq \sum_{s=t-K}^{t} s_{js} - I_{jt}, \quad j < J, \; t \in \mathcal{T}^{(2)},$$

$$x_{jt} \leq \sum_{s=t-K}^{t} s_{js} - I_{jt} + v_{jt}M, \quad j < J, \; t \in \mathcal{T}^{(2)},$$

$$x_{jt} \leq (1 - v_{jt})M, \quad j < J, \; t \in \mathcal{T}^{(2)},$$

where M is a preselected number larger than $\sum_{s=t-K}^{t} s_{js} - I_{jt}$ for all $j < J$ and $t \in \mathcal{T}^{(1)}$. Finally, we apply a similar approach to the delivery constraints by introducing additional binary variables $w_{1t} \in \{0, 1\}$, $t \in \mathcal{T}^{(1)}$. Using these variables we can transform the constraints (20) into the set of constraints

$$\begin{aligned}
&s_{1(t+l)} \leq d_t, \quad t \in \mathcal{T}^{(1)},\\
&s_{1(t+l)} \leq I_{1t} + s_{2(t+l)}, \quad t \in \mathcal{T}^{(1)},\\
&s_{1(t+l)} \geq d_t - w_{1t}M, \quad t \in \mathcal{T}^{(1)},\\
&s_{1(t+l)} \geq I_{1t} + s_{2(t+l)} - (1 - w_{1t})M, \quad t \in \mathcal{T}^{(1)}.
\end{aligned}$$

The constraints (21) can be dealt with in a similar way.

C Mixed integer linear formulation of service guarantee model

The reverse stress test model (2) for the service guarantee supply chain in Section 5.2 with policy (11), distance measure (15), and set of fault scenarios (14) is:

$$\min \quad \sum_{i=1}^{M}(c_i - \bar{c}_i)^2 + \sum_{j=1}^{N}(d_j - \bar{d}_j)^2 \tag{24}$$

$$\text{s.t.} \quad l_j^d = max\{d_j - \bar{s}_j - \sum_{i=1}^{M} \frac{d_j}{\bar{d}_j}\bar{x}_{ij}, 0\}, \qquad j = 1, \ldots, N, \tag{25}$$

$$l_i^c = max\{\sum_{j=1}^{N} \frac{d_j}{\bar{d}_j}\bar{x}_{ij} - c_i, 0\}, \qquad i = 1, \ldots, M, \tag{26}$$

$$\sum_{j=1}^{N} l_j^d(\mathbf{c}, \mathbf{d}) + \sum_{i=1}^{M} l_i^c(\mathbf{c}, \mathbf{d}) \geq \delta_2 \times \sum_{j=1}^{N} \bar{d}_j, \tag{27}$$

$$d_j \geq 0, \qquad j = 1, \ldots, N, \tag{28}$$

$$c_i \geq 0, \qquad i = 1, \ldots, M. \tag{29}$$

The objective function (24) minimizes the sum of euclidean distances between each capacity c_i and its baseline value $\bar{c}_i$ and each demand d_j and its baseline value $\bar{d}_j$. The goal is to find values for c and d that are as close as possible to their baseline values.

The constraints (25) calculate the unmet demand for each product j that is not covered by current inventory or production allocation. Using binary variables $w_j \in \{0, 1\}$, $j = 1, \ldots, N$ we can transform these constraints into the set of constraints

$$\begin{aligned}
&l_j^d \geq d_j - \bar{s}_j - \sum_{i=1}^{M} \frac{d_j}{\bar{d}_j}\bar{x}_{ij}, \qquad j = 1, \ldots, N,\\
&l_j^d \geq 0, \qquad j = 1, \ldots, N,\\
&l_j^d \leq d_j - \bar{s}_j - \sum_{i=1}^{M} \frac{d_j}{\bar{d}_j}\bar{x}_{ij} + w_j M, \qquad j = 1, \ldots, N,\\
&l_j^d \leq (1 - w_j)M, \qquad j = 1, \ldots, N.
\end{aligned}$$

The constraints (26) capture the excess demand in each plant i after accounting for the demand of all products assigned to that plant. Similarly, we can replace these constraints by the set of constraints

$$
\begin{aligned}
& l_i^c \geq \sum_{j=1}^{N} \frac{d_j}{\bar{d}_j} \bar{x}_{ij} - c_i, \qquad i = 1, \ldots, M, \\
& l_i^c \geq 0, \qquad i = 1, \ldots, M, \\
& l_i^c \leq \sum_{j=1}^{N} \frac{d_j}{\bar{d}_j} \bar{x}_{ij} - c_i + y_i M, \qquad i = 1, \ldots, M, \\
& l_i^c \leq (1 - y_i) M, \qquad i = 1, \ldots, M.
\end{aligned}
$$

The constraints (27) serves as a critical threshold for reverse stress testing, aiming to identify conditions that push the supply chain to a fault line, represented here by δ_2, which signals an unacceptable state.

The constraints (28)-(29) are non-negativity constraints.

D The SG model and baseline derivation

This appendix details the derivation of the baseline scenario. Specifically, we obtain the input parameters $\bar{d}_j$ and $\bar{c}_i$ from Simchi-Levi et al. (2018), and solve the SG model formulated below to compute $\bar{x}_{ij}$ and $\bar{s}_j$.

$$
\begin{aligned}
\min_{\bar{\mathbf{s}}, \bar{\mathbf{x}}, \mathbf{l}} \quad & \sum_{j=1}^{N} h_j \bar{s}_j \\
\text{s.t.} \quad & \sum_{i:(\mathcal{S}_i, \mathcal{T}_j) \in \mathcal{F}} \bar{x}_{ij} + l_j = \bar{d}_j - \bar{s}_j, \qquad j = 1, \ldots, N, \\
& \sum_{j:(\mathcal{S}_i, \mathcal{T}_j) \in \mathcal{F}} \bar{x}_{ij} \leq \bar{c}_i, \qquad i = 1, \ldots, M, \\
& \sum_{j=1}^{N} l_j \leq \delta_1, \\
& \bar{\mathbf{s}}, \bar{\mathbf{x}}, \mathbf{l} \geq 0.
\end{aligned}
\tag{30}
$$

In the original formulation of the first constraint, the inequality was written as $\geq$. However, this allowed x_{ij} values to exceed realistic levels, as the objective function did not penalize overproduction. To ensure the model better reflects practical scenarios and avoids overproduction, we modified the constraint to use equality, enforcing a strict balance between production and demand adjustments.

E Results of the 3-chain service guarantee model

Figure 9 shows the reverse stress testing results for the 3-chain design. Compared to the 2-chain design, the critical scenario was closer to the baseline, meaning smaller deviations could trigger shortages due to lower inventory levels per plant. This occurred because we first solved the SG model to determine inventory levels, which served as the baseline. Since the 3-chain design had more links, the resulting inventory levels were lower than in the 2-chain design.

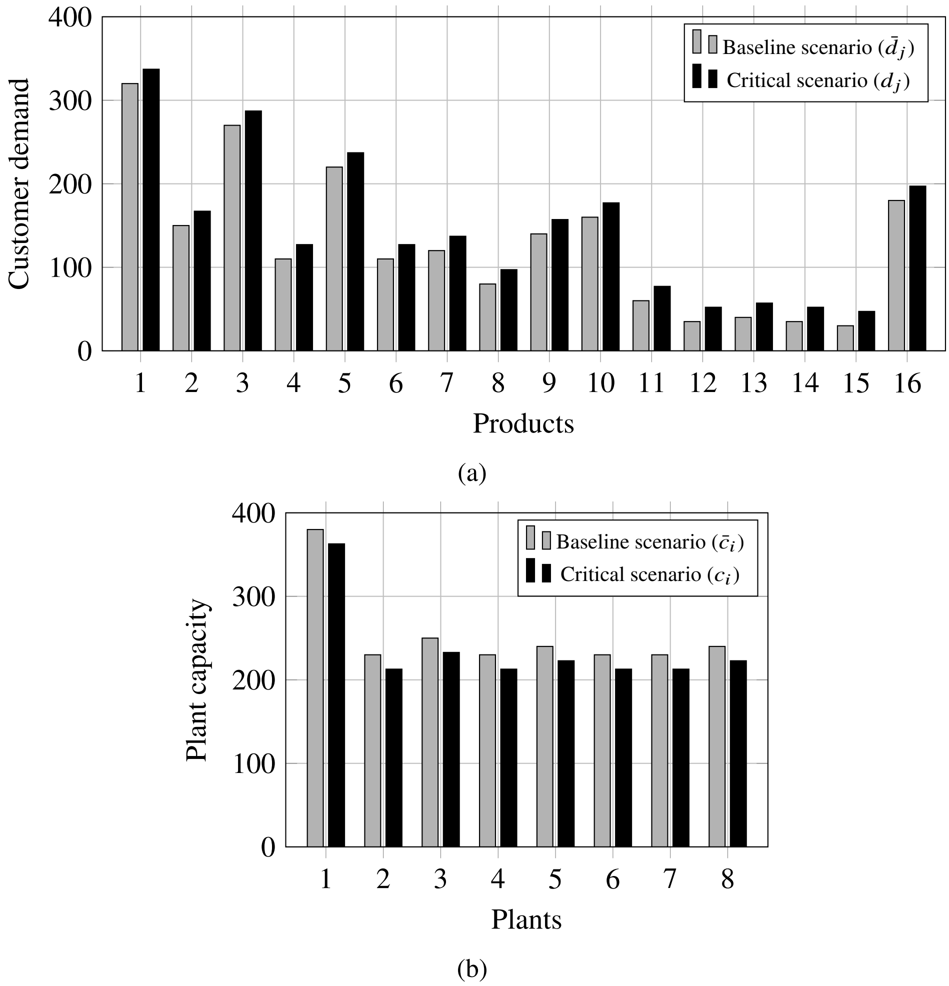


Figure 9: Deviations from the baseline for the 3-chain design: (a) product demand, (b) plant capacity.

F Additional results for smaller values of ϵ

To evaluate the effect of the neighborhood size on the diversity of the generated stress scenarios, we analyzed smaller values of ϵ based on the distances of the critical scenarios. Let Γ_1, Γ_2, and Γ_∞ denote the l_2 distances of the critical scenarios obtained from the l_1, l_2, and l_∞ distance measures, respectively (see Table 6). Based on these distances, we scaled the neighborhood bound $\Gamma_2(1 + \epsilon)$ to match Γ_1 and Γ_∞ to observe how tighter bounds restrict the resulting demand patterns. First, we determined the value of ϵ that satisfies the equation $\Gamma_2(1 + \epsilon) = \Gamma_1$, which results in $\epsilon = 0.34$. We then explored multiple solutions satisfying the inequality $\|d - \bar{d}\|_2 \leq \Gamma_2(1 + 0.34)$, using the l_1 critical scenario as the initial feasible solution. The generated multiple stress scenarios for this case are presented in Figure 10. Next, we expanded the search radius by finding the ϵ value that satisfies $\Gamma_2(1 + \epsilon) = \Gamma_\infty$, yielding $\epsilon = 0.77$. Subsequently, we searched for multiple solutions within the boundary $\|d - \bar{d}\|_2 \leq \Gamma_2(1 + 0.77)$, this time utilizing the l_∞ critical scenario as the initial solution. The resulting demand patterns for this relaxed bound are illustrated in Figure 11. As observed in both figures, limiting the neighborhood space with these smaller ϵ values restricts the model's feasible region, yielding significantly less diverse demand patterns compared to the broader exploration where $\epsilon = 1$ (as discussed in Section 5.1.4).

Distance	Value
Γ_1	6567254.02
Γ_2	4925193.23
Γ_∞	8713377.45

Table 6: The l_2 distances of the critical scenarios obtained under different distance measures.

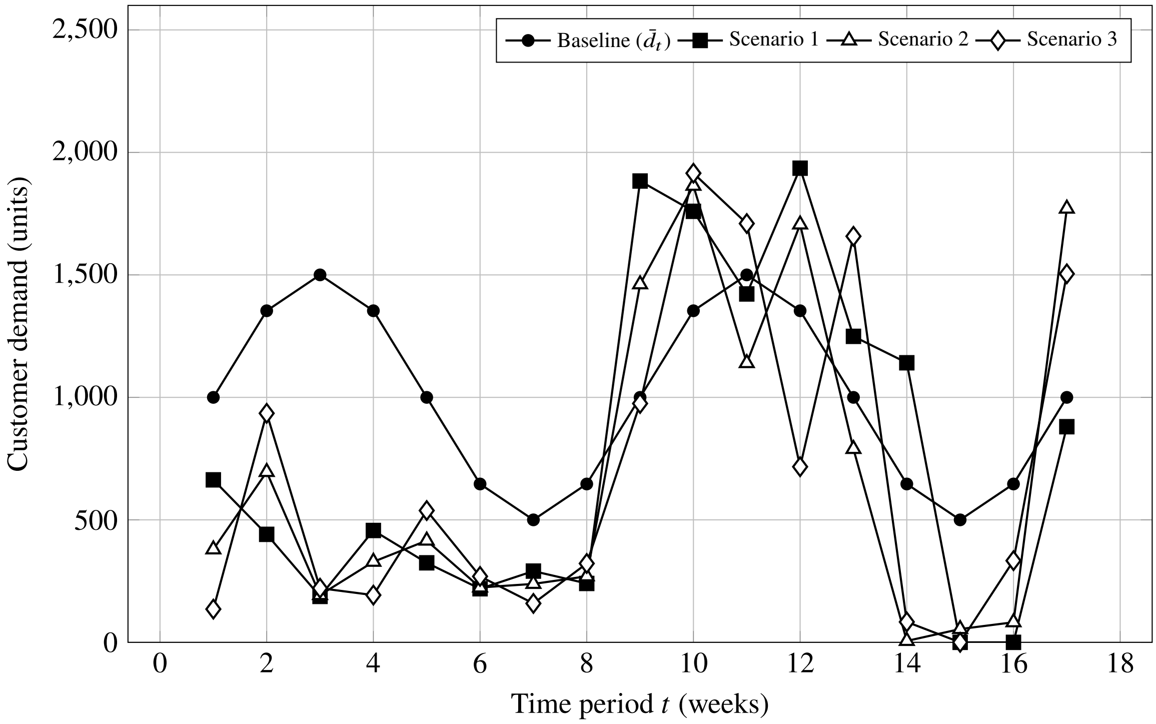


Figure 10: Baseline and multiple stress scenarios obtained from different random iterations (ϵ = 0.34).

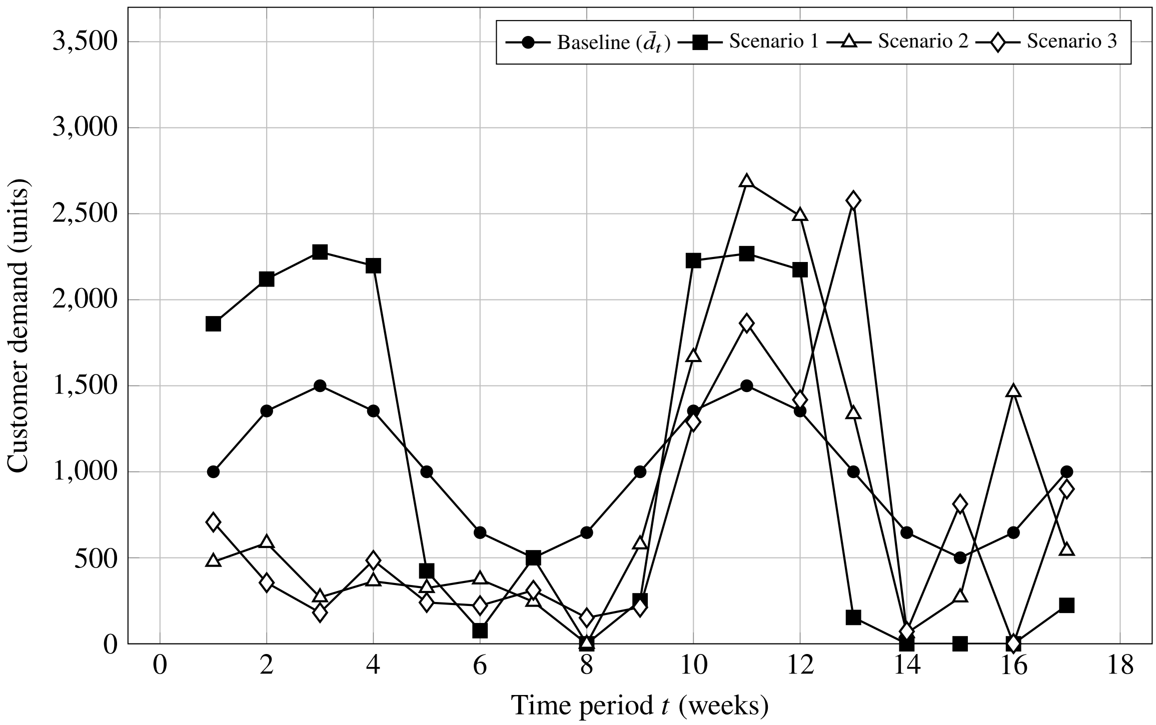


Figure 11: Baseline and multiple stress scenarios obtained from different random iterations (ϵ = 0.77).