INTERNATIONAL JOURNAL of ENGINEERING TECHNOLOGIES-IJET
Sarı Ay, Vol.10, No.4, 2025

Evaluating Strategic Workforce Decisions in Aggregate Production Planning under Demand Uncertainty: A Two-Stage Stochastic MILP with Out-of-Sample Assessment

Didem Sarı Ay*

* Assistant Professor, PhD., Department of Industrial Engineering, Alanya Alaaddin Keykubat University, Alanya, Antalya, Türkiye. E-mail: didem.ay@alanya.edu.tr ORCID: 0000-0002-9403-9015

Received: 01.02.2026 Accepted: 20.02.2026

Abstract - Aggregate production planning (APP) requires balancing workforce and operational decisions over a medium-term horizon. This study formulates and applies a two-stage stochastic mixed integer linear program (MILP) for APP with the primary objective of evaluating strategic workforce planning decisions under demand uncertainty. Workforce decisions are modeled as here-and-now commitments, while operational decisions are optimized as recourse actions in response to realized demand. The framework is demonstrated in an illustrative furniture manufacturing setting over a 12-month horizon with seasonally varying cost parameters. Demand scenarios are generated by combining Holt–Winters point forecasts with forecast-error scenarios obtained through a rolling-origin procedure and a moment-matching approach, yielding demand trajectories that reflect the statistical properties and temporal dependence of forecast uncertainty. Using these scenarios, the model quantifies cost–service trade-offs under alternative backorder penalty severities. To assess the robustness of the resulting workforce plans, this study conducts an out-of-sample evaluation based on observed demand from a holdout year and a wait-and-see benchmark, a validation perspective that has received limited attention in the APP literature. The out-of-sample results indicate that the stochastic model produces feasible and cost-effective workforce decisions that remain near-optimal under observed demand. Overall, the proposed framework serves as an effective decision-support tool for APP under demand uncertainty, supporting the evaluation of workforce and operational decisions within a unified stochastic framework.

Keywords: Aggregate production planning, Two-stage stochastic programming, Demand uncertainty, Scenario generation, Moment matching scenario generation, Out-of-sample evaluation

Talep Belirsizliği Altında Toplam Üretim Planlamasında Stratejik İşgücü Planlamasının Değerlendirilmesi: İki Aşamalı Stokastik MILP Yaklaşımı ve Örneklem Dışı Analizi

Öz

Toplam üretim planlaması, orta vadeli bir planlama ufkunda işgücü ve operasyonel kararlar arasında denge kurulmasını gerektirir. Bu çalışma, talep belirsizliği altında stratejik işgücü planlama kararlarını değerlendirmeyi amaçlayan iki aşamalı stokastik bir karma tamsayılı doğrusal programlama modeli geliştirmekte ve uygulamaktadır. İşgücü kararları, talep gerçekleşmeden önce verilen burada-ve-şimdi (here-and-now) kararlar olarak modellenirken, operasyonel kararlar gerçekleşen talebe bağlı olarak telafi (recourse) kararları şeklinde optimize edilmektedir. Önerilen çerçeve, mevsimsellik içeren maliyet yapısına sahip 12 aylık bir planlama ufku altında, örnek bir mobilya üretim uygulaması üzerinden gösterilmektedir.
Talep senaryoları, Holt–Winters nokta tahminleri ile kayan başlangıçlı (rolling-origin) bir prosedür ve moment eşleştirme yaklaşımı kullanılarak elde edilen tahmin hatası senaryolarının birleştirilmesiyle oluşturulmakta; böylece talep belirsizliğinin istatistiksel özelliklerini ve zamansal bağımlılığını yansıtan talep yörüngeleri elde edilmektedir. Bu senaryolar kullanılarak, farklı geri sipariş ceza şiddeti varsayımları altında maliyet–hizmet düzeyleri arasındaki ödünleşimler nicel olarak analiz edilmektedir. Elde edilen işgücü planlarını değerlendirmek amacıyla, bu çalışma ayrıca model kurulumunda kullanılmayan bir tutma yılına (holdout year) ait gözlemlenen talep verilerine dayalı bir örneklem dışı değerlendirme ve bekle-ve-gör (wait-and-see) kıyaslaması gerçekleştirmektedir. Toplam üretim planlama literatüründe sınırlı ölçüde ele alınan bu doğrulama yaklaşımı, stokastik planlamanın pratik performansına ilişkin nicel analizler sunmaktadır. Örneklem dışı sonuçlar, stokastik modelden elde edilen işgücü kararlarının gözlemlenen talep altında da uygulanabilir ve maliyet açısından etkin olduğunu, ayrıca mükemmel

bilgi varsayımı altındaki çözümlere yakın performans sergilediğini göstermektedir. Genel olarak, önerilen çerçeve talep belirsizliği altında toplam üretim planlaması için etkili bir karar destek aracı sunmakta ve işgücü ile operasyonel kararların bütünleşik bir stokastik yapı içinde değerlendirilmesine olanak sağlamaktadır.

Anahtar Kelimeler: Toplam üretim planlaması, İki aşamalı stokastik programlama, Talep belirsizliği, Moment eşleştirme tabanlı senaryo üretimi, Örneklem dışı değerlendirme

1. Introduction

Production planning is a critical decision-making problem in a wide range of applied domains, including manufacturing systems [1-4], supply chain management [5, 6], and production systems with integrated energy considerations [7, 8]. Aggregate production planning (APP) determines workforce levels, production quantities, inventory, overtime usage and backlog decisions over a medium-term (3-18 months) planning horizon. These decisions directly affect both operational costs and customer service performance. In practice, APP decisions must often be made in the presence of significant uncertainty. As a result, APP has been widely studied under uncertainty [9-12]. Sources of uncertainty may arise from customer demand [13], production costs [13, 14], processing times, available capacities , production quality, and other operational factors [15].

Among these sources, demand uncertainty is one of the most commonly studied in production planning due to its immediate impact on both the ability to satisfy customer demand and the efficient utilization of resources [16]. Fluctuations in demand may lead to either excess inventory or unmet customer demand if not properly anticipated. To address this challenge, stochastic programming has been widely applied in APP to explicitly model demand uncertainty and to evaluate alternative planning strategies under different scenario sets [1, 5, 10, 17]. In addition to demand uncertainty, production planning decisions are strongly influenced by managerial assumptions regarding backorder costs [10, 18]. Backorder costs represent penalties associated with delayed order fulfillment, such as loss of customer goodwill, expedited shipping, or contractual penalties. In practice, these costs are often difficult to estimate precisely and may vary depending on market conditions, customer expectations, and managerial priorities. Consequently, planners frequently rely on sensitivity analysis to understand how different backorder cost assumptions affect production and capacity decisions.

Two-stage stochastic programming [19] provides a natural framework for modeling such decision environments . In a two-stage structure, long-term decisions such as workforce levels are made before demand is realized, while short-term operational decisions such as production, inventory, overtime, and backlog are adjusted after demand becomes known [2, 10, 20-22]. This structure closely reflects real-world production planning processes with limited workforce flexibility and enables a systematic evaluation of cost and service trade-offs. In this study, the two-stage stochastic framework is used to evaluate strategic workforce planning decisions based on their expected operational performance across multiple demand scenarios, with the value of planning under uncertainty assessed through comparison with a wait-and-see benchmark representing perfect information.

The proposed stochastic APP model is formulated as a mixed integer linear program and demonstrated using an illustrative furniture manufacturing setting. Scenario-based probabilistic demand realizations are constructed from historical demand information, and backorder cost severity is examined through alternative penalty assumptions. This enables a systematic analysis of trade-offs among backlog, inventory, overtime, and total expected cost under different service-level priorities, using representative cost parameters that reflect realistic planning conditions.

While stochastic APP models are typically evaluated based on expected costs across demand scenarios, their performance under realized demand is often not explicitly examined. In practice, first-stage workforce decisions must be implemented before demand is observed, and their true quality depends on performance under actual demand realizations [23]. To address this issue, this study complements the scenario-based analysis with an out-of-sample evaluation using observed demand from a holdout year, comparing here-and-now solutions with a wait-and-see benchmark obtained under perfect information.

The main contributions of this study are threefold. First, the paper formulates a two-stage stochastic APP model that captures the sequential nature of workforce and operational decision making under demand uncertainty. This framework enables a systematic analysis of cost–service trade-offs across alternative backorder penalty severities at the aggregate planning level. Second, the study incorporates a forecast-based demand scenario generation approach that combines rolling-origin forecast errors with a moment-matching procedure, enabling the construction of demand scenarios that capture the key statistical properties and temporal dependence of forecast uncertainty over the planning horizon. Third, the paper complements the scenario-based analysis with an out-of-sample evaluation under observed demand, providing an explicit assessment of the robustness of stochastic workforce decisions in APP, a perspective that is not routinely examined in the APP literature. Together, these contributions demonstrate how scenario-based stochastic APP can serve as an effective decision-support tool for planning under demand uncertainty.

The remainder of the paper is organized as follows. Section 2 describes the two-stage stochastic modeling framework, the APP problem, and the mathematical formulation of the two-stage APP model. Section 3 presents an illustrative numerical application, discusses the computational results, analyzes the impact of backorder cost severity on production planning decisions, and assesses the performance of the stochastic program through an out-of-sample evaluation that compares here-and-now and wait-and-see solutions. Section 4 concludes the paper and outlines directions for future research.

2. Materials and Methods

2.1 Two-Stage Stochastic Modeling Framework

For minimization problems involving uncertain inputs, stochastic programming formulations often yield lower expected costs than deterministic models that rely solely on expected parameter values. This improvement arises from explicitly accounting for uncertainty in the decision-making process rather than replacing uncertain parameters with their mean values.

Two-stage stochastic programming extends this idea by distinguishing between here-and-now decisions that must be made before uncertainty is revealed and recourse decisions that are adjusted after uncertain parameters are realized. Two-stage stochastic programming has been widely applied to decision problems in which strategic or capacity-related decisions must be made before uncertainty is revealed, while operational decisions are adjusted afterward. Representative applications include facility and server location problems, where location and capacity decisions are made in advance and customer assignments depend on uncertain demand [24]; supply chain network design problems, in which investment and capacity decisions precede operational production and distribution decisions [25]. Two-stage formulations are also common in power systems applications such as unit commitment, where generation commitment decisions are made prior to demand realization and dispatch decisions are adjusted in real time [21, 26]. These applications share a common structure in which early-stage decisions constrain operational flexibility under uncertainty.

A formulation of a generic two-stage stochastic problem is [19]:

$$\min_{x} \; c^T x + E_{\xi} Q(x, \xi(\omega)) \tag{1}$$

$$s.t. \quad Ax = b, \tag{2}$$

$$x \in X, \tag{3}$$

where the recourse function is defined as

$$Q(x,\xi(\omega)) = \min_{y} \left\{ q(\omega)^T y \mid T(\omega)x + Wy = h(\omega), y \in Y \right\} \tag{4}$$

The first stage decisions variables x are chosen ex ante, before the realization of random events, $\omega \in \Omega$. The random vector, ξ, contains the stochastic parameters, $\xi(\omega) = \left(q(\omega), h(\omega), T(\omega)\right)$, affecting the second-stage problem. The second stage decisions, denoted by y, are determined once the realization of $\xi(\omega)$ is observed. The feasible regions for both stages, X and Y, may contain integrality constraints.

The objective of the two-stage stochastic program, given in Eq. (1), is to minimize the sum of the first-stage decision cost, $c^T x$, and the expected value of the optimal second-stage (recourse) cost. When the uncertain parameters are realized, the optimal recourse cost is computed by minimizing $Q(x,\xi(\omega))$, given in Eq. (4), with respect to y [19]. This modeling structure provides a flexible and realistic representation of decision environments in which some decisions must be made before uncertainty is resolved, while others can be adjusted after uncertainty becomes known.

2.2 Aggregate Production Planning Problem Description

The APP is formulated as a two-stage stochastic programming model. The first-stage decisions involve determining the workforce level in each period through hiring and firing actions before customer demand is realized. In practice, workforce planning decisions involve contractual, training, and organizational constraints, and are therefore treated as strategic decisions that are made in advance and fixed prior to demand realization, rather than adjusted monthly in response to observed demand. Once implemented, workforce decisions are fixed and must apply across all possible demand scenarios. The second stage decisions are operational and scenario-dependent. For each demand scenario, the model determines production quantities, overtime usage, inventory levels, and backlog amounts in each period. These recourse decisions enable the production system to respond to observed demand while respecting the workforce decisions made in the first stage.

The objective of the model is to minimize the total expected cost over the planning horizon: (i) first-stage workforce-related costs and (ii) the expected second-stage operational costs, including production, overtime, inventory holding, and backorder penalties, weighted by scenario probabilities. While a multi-stage formulation could be used to model gradual demand revelation, the two-stage structure adopted here reflects the limited flexibility of workforce decisions and provides a tractable yet realistic representation of strategic–operational interactions in aggregate planning.

2.3 Mathematical Formulation of the Two-Stage APP Model

This section presents the mathematical formulation of the two-stage stochastic MILP APP model described in Section 2. The notation, parameters, and decision variables used in the model are defined below.

Sets and Indices

$t \in \mathcal{T} = \{1,...,T\}$: time periods

$s \in S$: demand scenarios

Parameters

π_s: probability of scenario s, $\sum_{s \in S} \pi_s = 1$

D_t^s: demand in period t under scenario s

W_0: initial workforce

I_0: initial inventory

B_0: initial backlog

c^{reg}: regular capacity per worker per period
c^{ot}: maximum overtime capacity per worker per period
$W^{\min}$: minimum allowable workforce level
$\rho^{fire} \in [0,1]$: maximum fraction of workforce that can be laid off per period
$\alpha \in (0,1]$: capacity share
C_t^L: labor cost per worker
C_t^H: hiring cost per worker
C_t^F: firing cost per worker
C_t^P: regular production cost per unit
C_t^O: overtime production cost per unit
C_t^I: inventory holding cost per unit
C_t^B: backorder penalty per unit
$\bar{B}$: end-of horizon backlog cap

Decision Variables

First-stage decision variables
$W_t \in \mathbb{Z}_+$: workforce in period t
$H_t \in \mathbb{Z}_+$: number hired in period t
$F_t \in \mathbb{Z}_+$: number fired in period t

Second-stage decision variables
For each scenario s and period t:
$P_t^s \geq 0$: regular production
$O_t^s \geq 0$: overtime production
$I_t^s \geq 0$: end-of-period inventory
$B_t^s \geq 0$: end-of-period backlog

$\min \sum_{t \in T} \left(C_t^L W_t + C_t^H H_t + C_t^F F_t \right) + \sum_{s \in S} \pi_s \sum_{t \in T} \left(C_t^P P_t^s + C_t^O O_t^s + C_t^I I_t^s + C_t^B B_t^s \right)$	(5)
$W_t = W_{t-1} + H_t - F_t \quad \forall t$	(6)
$W_t \geq W^{\min} \quad \forall t$	(7)
$F_t \leq \rho^{\text{fire}} W_{t-1} \quad \forall t$	(8)
$I_t^s - B_t^s = I_{t-1}^s - B_{t-1}^s + P_t^s - D_t^s \quad \forall t,s$ with $I_0^s = I_0$ and $B_0^s = B_0$ given for all s.	(9)
$P_t^s \leq \alpha c^{\text{reg}} W_t + O_t^s \quad \forall t,s$	(10)
$O_t^s \leq \alpha c^{\text{ot}} W_t \quad \forall t,s$	(11)
$B_T^s \leq \bar{B} \quad \forall s$	(12)
$W_t, H_t, F_t \in \mathbb{Z}_+ \quad \forall t$	(13)
$P_t^s \geq 0,\ O_t^s \geq 0,\ I_t^s \geq 0,\ B_t^s \geq 0 \quad \forall t,s$	(14)

The objective function in Eq. (5) minimizes the sum of first-stage workforce adjustment costs and the expected second-stage operating costs. Eq. (6) describes workforce evolution across periods in the first stage, while Eq. (7) and Eq. (8) enforce the minimum workforce requirement and the maximum allowable layoff rate, respectively. Eq. (9) represents the inventory balance in the second stage. For each period and scenario, inventory and backlog levels are updated based on previous-period values, production decisions, and realized demand, while accounting for the given initial inventory and backlog conditions.

3. Numerical Applications

The APP framework is demonstrated using an illustrative mid-sized furniture manufacturing setting. The analysis is based on constructed demand and cost parameters designed to reflect realistic planning conditions for medium-scale manufacturers. In the literature, similar studies employ illustrative or fictional applications presented through

Eq. (10) limits regular production by the available workforce capacity, linking production decisions to workforce levels through the regular production rate. Eq. (11) restricts overtime production, reflecting operational limits on overtime usage. Eq. (12) imposes an upper bound on the backlog level in the final period T, preventing excessive unmet demand at the end of the planning horizon. Eq. (13) enforces integrality requirements on workforce-related decision variables, and Eq. (14) ensures the nonnegativity of all production, inventory, backlog, and overtime variables. The resulting model is a MILP and is solved to optimality using IBM ILOG CPLEX numerical examples to capture a range of situations that may arise in industrial practice [27, 28]

The numerical application focuses on a single product family: living room sets. A living room set is treated as a single aggregate product and includes items such as sofas, armchairs, and television stands. To reflect the shared use of production resources across multiple product families, a capacity share parameter is introduced, with 40% of the total workforce capacity allocated to living room set production.

The planning horizon consists of 12 monthly periods. Workforce decisions, including hiring and layoffs, are made prior to demand realization and are modeled as first-stage decisions. Production, overtime, inventory, and backlog decisions are made after demand is realized and are modeled as second-stage recourse decisions.

Demand is constructed based on existing studies in the literature that report monthly demand data for furniture manufacturing settings [3]. In constructing the demand data, key characteristics observed in practice—such as seasonality, temporal interdependencies, and structural demand shifts—are taken into account. The resulting synthetic demand series incorporates a baseline level with mild growth, time-varying seasonal patterns, and periods of heightened uncertainty to reflect realistic demand trajectories. In reference [27], Gnanendran et al. construct demand values by drawing from a continuous uniform distribution. Demand uncertainty is then represented using probabilistic scenarios derived from these constructed historical demand patterns capturing the key statistical properties and temporal dependence of demand over the planning horizon.

The planning horizon is assumed to begin with an initial workforce of 50 workers. Production capacity is modeled as a function of workforce size, and overtime is permitted within realistic operational limits. Additional constraints are included to prevent excessive workforce reductions and to ensure operational continuity.

Cost parameters are selected to be plausible and internally consistent with typical manufacturing cost structures. They are intended to support comparative analysis across backorder-cost assumptions rather than to replicate a specific firm's accounting data.

3.1 Parameter Settings and Demand Scenarios

The parameter values used in the numerical application of the model are summarized in Table 1, along with their descriptions.

Table 1. Parameter settings and cost structure used in the numerical application.

Parameter	Value	Units	Description
T	12	month	Number of planning periods
W_0	50	worker	Initial workforce
I_0	80	units	Initial Inventory
B_0	0	units	Initial backorder
c^{reg}	10	units/worker/month	Regular production capacity per worker
c^{ot}	2	units/worker/month	Maximum overtime capacity per worker
$W^{\min}$	20	worker	Minimum workforce level
ρ^{fire}	0.10	fraction	Maximum fraction of workforce that can be laid off per period
α	0.40	fraction	Capacity share allocated to living room sets
C_0^L	40,000	TRY/worker/month	Labor cost
C_0^H	50,000	TRY/worker	Hiring cost
C_0^F	80,000	TRY/worker	Layoff cost
C_0^P	4,500	TRY/unit	Regular production cost
C_0^O	6,750	TRY/unit	Overtime production cost
C_0^I	100	TRY/unit/month	Inventory holding cost
C_t^B	scenario-dep.	TRY/unit/month	Backorder penalty cost
$\bar{B}$	10	units	End-of-horizon backlog cap
μ_t^{LP}	$(1,1.03,1.06,1.10)$	quarterly	Seasonal multiplier for labor, hiring, and regular production costs
μ_t^{OT}	$(1,1.05,1.12,1.20)$	quarterly	Seasonal multiplier for overtime production cost
μ_t^{I}	$(1,1.01,1.03,1.05)$	quarterly	Seasonal multiplier for inventory holding cost
μ_t^{B}	$(1,1.05,1.15,1.25)$	quarterly	Seasonal multiplier for backorder penalty cost

We applied a forecast-based demand scenario generation approach for the next year. Probabilistic demand scenarios are constructed to capture plausible realizations of monthly demand for the upcoming planning year, conditional on the information available at the time of planning. A ten-year dataset of historical monthly demand observations for living room sets is used to estimate demand trends, seasonal patterns, and forecast uncertainty. Demand forecasts are generated using the Holt–Winters exponential smoothing method with additive trend and additive seasonality, which has been widely applied for forecasting time series with trend and seasonal components [29]. To obtain a realistic representation of forecast uncertainty, a rolling-origin evaluation procedure is employed. Specifically, for each year $y = 2,...,10,$ monthly demand forecasts for year y are generated using only demand observations from years $1,...,y-1.$ Let $d_{m,y}$ denote the observed demand in month m of year y, and let $\hat{d}_{m,y}$ denote the corresponding forecast obtained from this information set. The forecast error is then defined as

$$e_{m,y} = d_{m,y} - \hat{d}_{m,y}, \quad y = 2,...,10, \ m = 1,...,12. \tag{15}$$

Using all available demand observations from years $1,...,10,$ a point forecast $\hat{d}_{m,11}$ is generated for each month $m = 1,...,12$ of the next planning year (Year 11). Uncertainty around this point forecast is represented through forecast-error scenarios generated via a moment-matching procedure [30, 31]. Historical forecast errors obtained from the rolling-origin procedure are used to estimate the first and second moments of monthly demand uncertainty. In particular, the mean vector and variance–covariance matrix of the forecast error process are computed, capturing both the marginal variability of monthly errors and their temporal dependence across months. A set of multivariate forecast-error scenarios is then generated such that the sample moments of the scenario set match these estimated moments.

Let $e^{(s)}_{m,11}$ denote the forecast error in month m of Year 11 under scenario $s.$ Each demand scenario for the next planning year is then constructed by adding the forecast-error realization to the corresponding point forecast:

$$d^{(s)}_{m,11} = \hat{d}_{m,11} + e^{(s)}_{m,11}, \quad \forall m. \tag{16}$$

Nonnegativity of demand is enforced by truncation if necessary. All scenarios are assigned equal probability, $p_s = 1/S.$ Using this procedure, $S = 20$ demand scenarios are generated and used as inputs to the stochastic APP model. The resulting scenarios preserve key statistical properties of forecast uncertainty while providing realistic next-year demand trajectories, which are illustrated in Fig. 1.

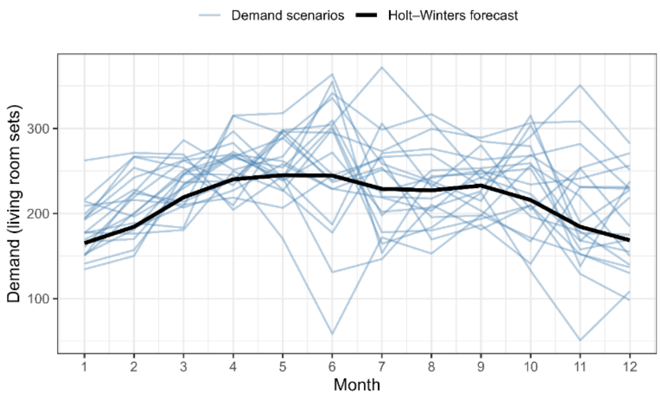

Fig. 1. Monthly demand scenarios

3.2 Numerical Results and Backorder Cost Sensitivity Analysis

In this section, we summarize the numerical results and analyze how varying backorder cost assumptions affect APP decisions. Similar trade-off analyses have been explored in the production planning literature [16, 32, 33] to evaluate the performance of APP strategies.

All computational experiments are conducted in Python (version 3.13.1) using the PuLP optimization library and solved to optimality with IBM ILOG CPLEX.

Tables 2, 3, and 4 report the results obtained under three backorder cost severity levels when seasonal cost variations are incorporated into the APP model. Sensitivity to shortage or backorder penalties has been explored in the APP literature through alternative penalty structures—such as breakpoint-based shortage cost formulations—to examine how aggregate plans shift between backlog accumulation and the use of capacity or inventory buffers [16, 17].

Table 2. Cost and inventory outcomes by backorder severity.

Case	Total expected cost	Total expected inventory	Max expected inventory (monthly)	Max inventory (any scenario/month)
Low backorder severity	36,648,008	798.00	143.20	261
Medium backorder severity	36,772,119	944.60	166.25	273
High backorder severity	37,123,613	1,059.75	163.75	293

Table 3. Backlog and overtime outcomes by backorder severity.

Case	Total expected backlog	Max expected backlog (monthly)	Total expected overtime units
Low backorder severity	131.75	24.85	129.25
Medium backorder severity	93.80	15.80	135.00
High backorder severity	39.75	9.05	148.95

Table 4. Workforce response metrics by backorder severity.

Case	Average workforce	Workforce variability (L1 norm)	Total layoffs (fires)
Low backorder severity	48.0	0	2
Medium backorder severity	47.8	3	3
High backorder severity	47.0	4	4

When backorder severity is low, the model allows a larger amount of unmet demand. The expected total backlog (131.75 units) and the maximum expected monthly backlog (24.85 units) are highest in this case. Workforce decisions remain largely unchanged over the planning horizon, with no hiring and only limited layoffs, while overtime usage remains moderate.

As backorder severity increases to the medium level, backlog levels decline substantially. This improvement in service performance is primarily achieved through increased use of overtime, while workforce adjustments remain limited and the average workforce decreases only slightly over the horizon. Under high backorder severity, backlog becomes increasingly costly and is therefore reduced aggressively. To achieve this higher service level, the model relies more heavily on overtime usage and workforce adjustments, resulting in higher total expected costs. Overall, these results clearly illustrate the fundamental cost–service trade-off in APP under demand uncertainty.

Fig. 2 presents the expected inventory, backlog, and overtime levels over the planning horizon. When backorder severity is low, the model allows backlog to accumulate primarily in the middle and later periods of the horizon, while inventory levels decline steadily and approach zero toward the end of the planning period. In this case, overtime usage remains limited, indicating that the model accepts delayed demand fulfillment as a cost-effective option when backorders are inexpensive, even in the presence of seasonally increasing production and labor costs.

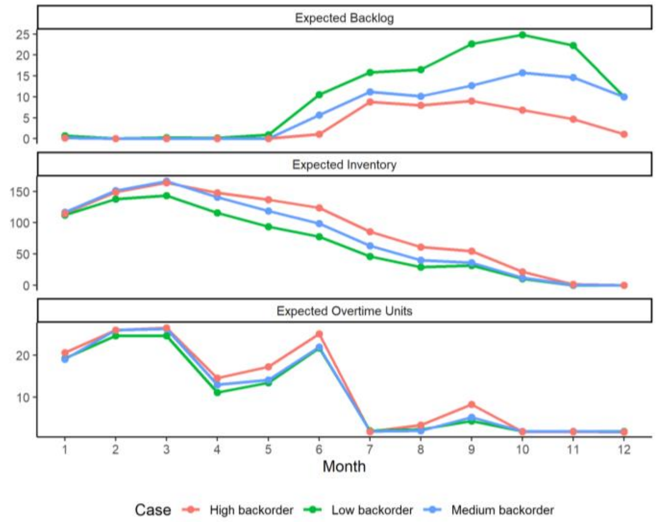

Fig. 2. Expected inventory, backlog, and overtime levels by backorder severity.

As backorder severity increases, backlog levels are noticeably reduced, particularly in the later months of the horizon. This reduction is achieved through greater reliance on overtime production and the maintenance of higher inventory levels, especially during the early and mid-horizon periods when seasonal cost pressures are lower. Under high backorder severity, the model builds inventory buffers earlier in the planning horizon and utilizes overtime capacity more intensively, resulting in consistently lower backlog levels across all periods. Overall, these results illustrate how stricter service-level requirements, combined with seasonal cost escalation, shift the balance from backlog tolerance toward proactive capacity utilization.

Fig. 3 highlights the monthly deviations of the low and high backorder severity cases relative to the medium backorder severity benchmark, making the underlying policy differences more explicit.

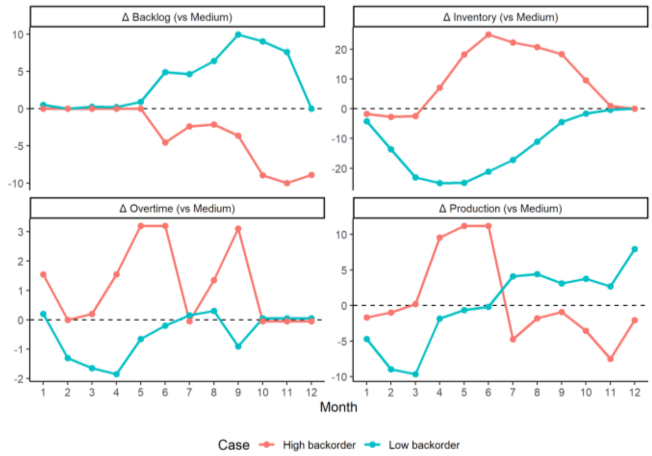

Fig. 3. Monthly deviations from the medium backorder case

Compared to the medium case, the low backorder severity scenario exhibits predominantly positive backlog deviations and negative inventory deviations across most months, indicating a backlog-oriented strategy in which delayed demand fulfillment is preferred over inventory buffering or capacity expansion. Overtime deviations in this case remain generally negative or close to zero, reinforcing the limited use of capacity-intensive adjustments.

In contrast, the high backorder severity scenario displays consistently negative backlog deviations, accompanied by positive inventory and overtime deviations, particularly in the early and middle portions of the planning horizon. This pattern indicates that, when backorders are heavily penalized, the model substitutes backlog with inventory buffering and increased capacity utilization, even in the presence of seasonally increasing cost conditions. Production deviations remain relatively modest, suggesting that the primary operational response to higher backorder penalties is achieved through inventory accumulation and overtime usage rather than substantial changes in regular production levels.

Overall, higher backorder penalties shift the aggregate production plan from backlog tolerance toward improved service performance through higher inventory buffers and greater reliance on capacity-intensive decisions such as overtime. Under seasonal cost escalation, this shift leads to earlier inventory buildup to mitigate costly backlog accumulation in later periods. Although these strategies increase total expected cost, they substantially improve service outcomes, highlighting the fundamental cost–service trade-off in APP under demand uncertainty.

3.3 Out-of-Sample Evaluation under Observed Demand

To evaluate the performance of the stochastic APP model under observed conditions, an out-of-sample evaluation [23] is conducted using monthly demand observed in a holdout year that is excluded from scenario construction. Several studies assess stochastic production plans against deterministic or alternative benchmark policies through simulation-based assessments under demand uncertainty [2, 12, 21, 34, 35]. While out-of-sample evaluations are conducted in references [21, 36-38], they are largely applied in alternative stochastic optimization settings (e.g., unit commitment, assemble-to-order production systems), and their use for evaluating APP decisions under observed demand is relatively limited. The out-of-sample evaluation presented in this study therefore provides additional insight into the robustness and practical effectiveness of stochastic APP under demand uncertainty.

For each backorder cost severity case, the first-stage workforce decisions—including workforce levels, hiring, and layoffs—obtained from the two-stage stochastic model are treated as fixed. Conditional on these fixed workforce decisions, the second-stage recourse problem—comprising regular production, overtime, inventory, and backlog decisions—is re-optimized under the observed demand trajectory. This procedure yields the realized operating cost and the corresponding total realized cost associated with each planning policy. Mathematically, this evaluation corresponds to solving the second-stage recourse problem with the first-stage decisions fixed at their here-and-now values and uncertainty replaced by the observed demand trajectory.

This analysis addresses the following question: *If this demand trajectory had occurred, what would have been the minimum achievable operating cost given the fixed workforce plan?* In this way, the out-of-sample evaluation provides an ex post assessment of the quality of the first-stage decisions - namely, workforce decisions made in advance- obtained from the two-stage stochastic program [21]. Let $\hat{\xi}$ denote the observed realization of the uncertain parameters (e.g., the holdout-year demand trajectory). Let x^{HN} denote the here-and-now solution obtained from solving Eq. (1)–(4) (e.g., workforce plan). We re-optimize the second-stage recourse problem under observed uncertainty. The observed recourse cost is

$$Q(x^{HN},\hat{\xi}) = \min_{y}\left\{ q\left(\hat{\xi}\right)^{\mathrm{T}} y \mid T\left(\hat{\xi}\right)x^{HN} + Wy = h\left(\hat{\xi}\right), y \in Y \right\}. \quad (17)$$

Then, the observed second-stage (operating) cost and the observed total cost (workforce cost and observed operating cost) are

$$C_{\text{op}}^{\text{real}} = Q(x^{HN},\hat{\xi}) \quad (18)$$

$$C_{\text{total}}^{\text{real}} = c^{\mathrm{T}} x^{HN} + Q(x^{HN},\hat{\xi}). \quad (19)$$

In the numerical application, the evaluation is based on historical monthly demand data spanning eleven years. Demand observations from the first ten years are used to construct probabilistic demand scenarios for a 12-month planning horizon and to compute the here-and-now workforce plans by solving the two-stage stochastic APP model. These workforce decisions are then implemented under the observed demand realized in the holdout year, which is excluded from model construction. Then, the corresponding second-stage recourse problem is solved to evaluate realized system outcomes.

Table 4 summarizes the results of the out-of-sample evaluation. For each backorder cost severity case, the table reports the minimum feasible operating cost consistent with the fixed workforce plan under the holdout-year demand realization, together with the associated inventory, backlog, and overtime outcomes. These realized costs represent a lower bound on the operating cost attainable under each fixed workforce plan and are used solely for comparative assessment across backorder cost regimes.

Table 4. Out-of-sample cost and operational outcomes under observed demand.

	Costs			Operational outcomes		
Backorder severity	First-stage	Sec-stage	Total	Inventory	Backlog	Overtime
Low	24,294,400	14,101,900	38,396,300	1,033	20	267
Medium	24,355,200	14,121,287	38,476,487	1,250	16	267
High	24,423,200	14,180,398	38,603,598	1,382	8	269

Across all cases, realized backlog decreases and inventory and overtime usage increase as backorder cost severity rises, indicating that the cost–service trade-off observed in the in-sample stochastic results persists under observed demand. Moreover, all workforce plans remain feasible and yield comparable cost levels, suggesting that the first-stage decisions obtained from the stochastic model are robust to out-of-sample demand realizations.

As a benchmark, a wait-and-see solution is obtained by solving a deterministic APP model under perfect information, using the full observed demand sequence of the holdout year [12, 34].

$$C_{\text{WS}}^{\text{real}}\left(\hat{\xi}\right) = \min_{x \in X}\left\{ c^{\mathrm{T}} x + Q\left(x,\hat{\xi}\right) \mid Ax = b \right\}. \quad (20)$$

Comparing the realized total cost of the here-and-now solution with that of the wait-and-see benchmark provides an out-of-sample measure of the economic value associated with perfect demand foresight at the workforce-planning stage for the holdout year.

$$\Delta^{\text{real}}\left(\hat{\xi}\right) = c^{\mathrm{T}} x^{HN} + Q(x^{HN},\hat{\xi}) - C_{\text{WS}}^{\text{real}}\left(\hat{\xi}\right). \quad (21)$$

In Ref [12], Jamalnia et al. applied a wait-and-see approach to evaluate the performance of APP strategies where they adopted a nonlinear multi-objective optimization model.

3.3.1 Cost Decomposition under Fixed Workforce Decisions

Building on the out-of-sample results summarized in Table 4, this subsection examines how the realized total cost decomposes into first-stage and second-stage components under fixed workforce decisions. Figure 4 compares the first-stage and second-stage costs associated with the here-and-now and wait-and-see solutions across different backorder cost severity levels. This decomposition provides insight into how perfect demand information affects workforce-related costs and operational recourse costs under observed demand conditions.

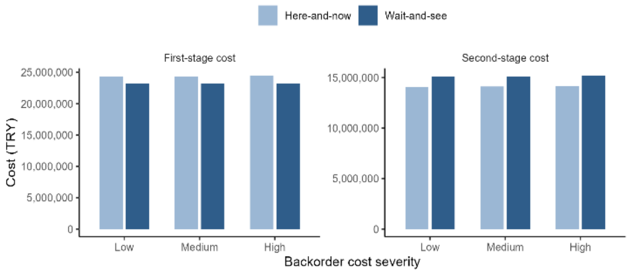

Fig. 4. Comparison of first-stage and second-stage costs for here-and-now and wait-and-see solutions under varying backorder cost severity levels.

Across all cases, the here-and-now solutions exhibit higher first-stage costs than the wait-and-see benchmark, reflecting more conservative workforce commitments made to hedge against demand uncertainty. These higher workforce-related costs represent an insurance effect rather than inefficiency. In contrast, the wait-and-see solutions incur higher second-stage costs, as lower workforce commitments under perfect information shift a greater share of demand adjustment to operational decisions such as overtime usage, inventory adjustments, and backlog management.

This cost decomposition highlights the fundamental trade-off between advance workforce planning under uncertainty and operational flexibility after demand realization, and provides a basis for understanding the operational adjustment mechanisms discussed next.

3.3.2 Operational Implications and Value of Perfect Information

While the cost decomposition in Section 3.3.1 clarifies how total costs differ between the here-and-now and wait-and-see solutions, the associated second-stage outcomes provide insight into the operational mechanisms through which these cost differences arise. Figure 5 compares total inventory, backlog, and overtime usage under the here-and-now and wait-and-see solutions across backorder cost severity levels for the holdout-year demand realization.

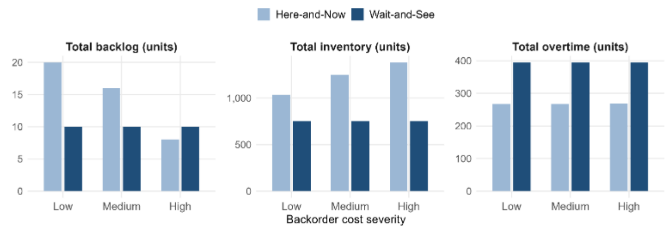

Fig. 5. Comparison of second-stage operational outcomes under here-and-now and wait-and-see solutions across backorder cost severity levels

As shown in Fig. 5, the operational outcomes under the here-and-now and wait-and-see solutions are consistent with the cost decomposition results in Section 3.3.1. Relative to the wait-and-see benchmark, the here-and-now solutions rely more on inventory buffering and less on overtime usage, particularly under higher backorder cost severity, reflecting precautionary capacity allocation under demand uncertainty. In contrast, the wait-and-see solutions accommodate demand fluctuations primarily through overtime adjustments under perfect information. These results confirm that the cost differences between the two approaches arise from distinct operational adjustment mechanisms rather than infeasibility or modeling artifacts.

Fig. 6 reports the realized cost gap between the here-and-now and wait-and-see solutions, expressed as a percentage of the wait-and-see first-stage cost. This normalized gap provides an economic measure of the cost of imperfect information at the workforce-planning stage. Because the here-and-now solution is evaluated using optimal second-stage recourse decisions, the observed percentage difference primarily reflects the value of being able to select workforce decisions with perfect information, rather than inefficiencies in operational decision making.

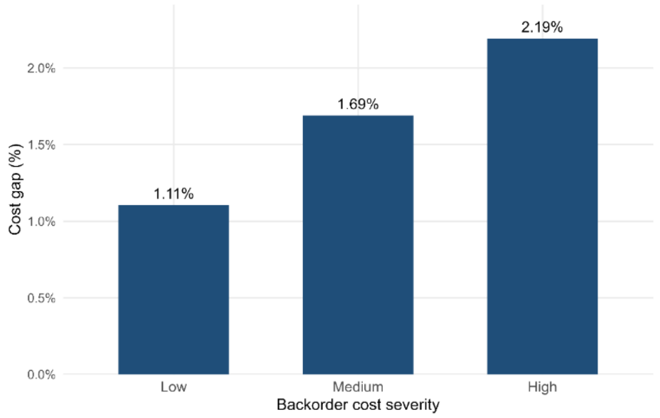

Fig. 6. Impact of backorder cost severity on the realized cost advantage of perfect information under observed demand.

The results show that the higher backorder penalties increase the value of perfect information at the workforce-planning stage. As shown in Fig. 6, the cost gap remains below 3% across all cases, indicating that the stochastic workforce plans remain economically robust under demand uncertainty.

Overall, the out-of-sample evaluation confirms that the two-stage stochastic APP model produces first-stage workforce decisions that remain feasible and cost-effective under observed demand, with economically interpretable deviations from the perfect-information benchmark.

4. Conclusions

This study examines APP decisions under demand uncertainty using a two-stage stochastic MILP framework, with a particular focus on evaluating strategic workforce decisions. Through an illustrative furniture manufacturing application, the analysis shows that assumptions regarding backorder cost severity play a central role in shaping aggregate production plans. Higher backorder penalties lead to improved service performance but require increased workforce capacity, overtime usage, and associated costs, highlighting clear cost–service trade-offs at the aggregate planning level.

An out-of-sample evaluation based on observed demand demonstrates the robustness of the stochastic planning framework. The first-stage workforce decisions obtained from the stochastic model remain feasible and cost-effective when confronted with unseen demand realizations, with realized costs remaining close to a perfect-information benchmark. Compared with highly sensitive stochastic planning problems such as two-stage unit commitment in electricity systems [26], the aggregate nature of APP and the availability of operational recourse options contribute to relatively robust workforce decisions under demand uncertainty.

From a practical perspective, the proposed framework serves as a decision-support tool that enables managers to evaluate the cost and service implications of alternative backorder severity assumptions under time-varying cost conditions. By adjusting a limited set of parameters, decision makers can explore alternative workforce and capacity strategies, evaluate their robustness to demand uncertainty, and better align planning decisions with operational priorities.

Overall, the results demonstrate that two-stage stochastic APP provides a practical and effective framework for workforce and production planning under uncertainty. Future research may consider rolling-horizon implementations, richer demand information structures, or extensions that incorporate additional sources of uncertainty, such as product quality and return flows.

References

[1] J. E. Gómez-Rocha, E. S. Hernández-Gress, and H. Rivera-Gómez, "Production planning of a furniture manufacturing company with random demand and production capacity using stochastic programming," (in English), *Plos One,* vol. 16, no. 6, Jun 14 2021, doi: ARTN e025280110.1371/journal.pone.0252801.

[2] Z. Y. Hu and G. P. Hu, "A two-stage stochastic programming model for lot-sizing and scheduling under uncertainty," (in English), *Int J Prod Econ,* vol. 180, pp. 198–207, Oct 2016, doi: 10.1016/j.ijpe.2016.07.027.

[3] F. J. D. Lina Gozali, Andri Tan Wijaya, "Production Planning and Control in Furniture Company at PT. Lion Metal Works," in *Proceedings of the 11th Annual International Conference on Industrial Engineering and Operations Management* Singapore, 2021.

[4] V. F. Yu, H. C. Kao, F. Y. Chiang, and S. W. Lin, "Solving Aggregate Production Planning Problems: An Extended TOPSIS Approach," (in English), *Appl Sci-Basel,* vol. 12, no. 14, Jul 2022, doi: ARTN 694510.3390/app12146945.

[5] S. C. H. Leung, Y. Wu, and K. K. Lai, "A stochastic programming approach for multi-site aggregate production planning," (in English), *J Oper Res Soc,* vol. 57, no. 2, pp. 123–132, Feb 2006, doi: 10.1057/palgrave.jors.2601988.

[6] M. Schön, C. H. Dai, F. Herrmann, and T. Claus, "Resilient Multi-Site Aggregate Production Planning: A Stochastic Model," (in English), *Commun Ecms,* vol. 39, pp. 434–440, 2025. [Online]. Available: <Go to ISI>://WOS:001572352000057.

[7] A. Karimi-Zare, G. H. Shakouri, A. Kazemi, and E. S. Kim, "Aggregate production planning and energy supply management in steel industry with an onsite energy generation system: A multi-objective robust optimization model," (in English), *Int J Prod Econ,* vol. 269, Mar 2024, doi: ARTN 10914910.1016/j.ijpe.2024.109149.

[8] M. Modarres and E. Izadpanahi, "Aggregate production planning by focusing on energy saving: A robust optimization approach," (in English), *J Clean Prod,* vol. 133, pp. 1074–1085, Oct 1 2016, doi: 10.1016/j.jclepro.2016.05.133.

[9] M. Qasim, K. Y. Wong, and Komarudin, "A review on aggregate production planning under uncertainty: Insights from a fuzzy programming perspective," (in English), *Eng Appl Artif Intel,* vol. 128, Feb 2024, doi: ARTN 10743610.1016/j.engappai.2023.107436.

[10] S. M. J. M. Al-e-Hashem, A. Baboli, and Z. Sazvar, "A stochastic aggregate production planning model in a green supply chain: Considering flexible lead times, nonlinear purchase and shortage cost functions," (in English), *Eur J Oper Res,* vol. 230, no. 1, pp. 26–41, Oct 1 2013, doi: 10.1016/j.ejor.2013.03.033.

[11] A. Jamalnia, J. B. Yang, A. Feili, D. L. Xu, and G. Jamali, "Aggregate production planning under uncertainty: a comprehensive literature survey and future research directions (vol 102, pg 159, 2019)," (in English), *Int J Adv Manuf Tech,* vol. 102, no. 1-4, pp. 183–183, May 2019, doi: 10.1007/s00170-019-03318-9.

[12] A. Jamalnia, J. B. Yang, D. L. Xu, A. Feili, and G. Jamali, "Evaluating the performance of aggregate production planning strategies under uncertainty in soft drink industry," (in English), *J Manuf Syst,* vol. 50, pp. 146–162, Jan 2019, doi: 10.1016/j.jmsy.2018.12.009.

[13] M. Doodman, F. Soleimani, F. Jolai, and A. Aghsami, "An integrated scenario-based multi-period multi-objective mathematical model for a sustainable biomass supply chain network considering social factors and disruption risk under uncertainty," (in English), *Environ Dev Sustain,* Jul 17 2024, doi: 10.1007/s10668-024-05234-9.

[14] U. M. Modibbo, S. Gupta, A. Ahmed, and I. Ali, "An integrated multi-objective multi-product inventory managed production planning problem under uncertain environment," (in English), *Ann Oper Res,* vol. 339, no. 3, pp. 1679–1723, Aug 2024, doi: 10.1007/s10479-022-04795-0.

[15] M. G. Mirzaei, S. Gholami, and D. Rahmani, "A mathematical model for the optimization of agricultural supply chain under uncertain environmental and financial conditions: the case study of fresh date fruit," (in

English), *Environ Dev Sustain,* vol. 26, no. 8, pp. 20807–20840, Aug 2024, doi: 10.1007/s10668-023-03503-7.

[16] G. Zhang, J. Shang, and W. Li, "Collaborative production planning of supply chain under price and demand uncertainty," *Eur J Oper Res,* vol. 215, no. 3, pp. 590–603, 2011/12/16/ 2011, doi: https://doi.org/10.1016/j.ejor.2011.07.007.

[17] J. Biazzi, "Aggregate planning for probabilistic demand with internal and external storage," *Journal of Operations and Supply Chain Management,* vol. 11, p. 37, 06/15 2018, doi: 10.12660/joscmv11n1p37-52.

[18] D. Rahmani, R. Ramezanian, P. Fattahi, and M. Heydari, "A robust optimization model for multi-product two-stage capacitated production planning under uncertainty," (in English), *Appl Math Model,* vol. 37, no. 20-21, pp. 8957–8971, Nov 1 2013, doi: 10.1016/j.apm.2013.04.016.

[19] J. R. Birge and F. Louveaux, "Introduction to Stochastic Programming, Second Edition," (in English), *Springer Ser Oper Re,* pp. 3–+, 2011, doi: 10.1007/978-1-4614-0237-4.

[20] S. C. H. Leung, & Ng, W. L., "A stochastic programming model for production planning of perishable products with postponement," *Production Planning & Control - PRODUCTION PLANNING CONTROL,* vol. 18(3), 190–202, 2007, doi: https://doi.org/10.1080/09537280601009047.

[21] D. S. Ay and S. M. Ryan, "Observational data-based quality assessment of scenario generation for stochastic programs," (in English), *Comput Manag Sci,* vol. 16, no. 3, pp. 521–540, Jul 2019, doi: 10.1007/s10287-019-00349-1.

[22] S. R. Islam, C. Novoa, and T. D. Jin, "Multi-facility aggregate production planning with prosumer microgrid: A two-stage stochastic program," (in English), *J Clean Prod,* vol. 367, Sep 20 2022, doi: ARTN 13291110.1016/j.jclepro.2022.132911.

[23] M. Kaut and S. W. Wallace, *Evaluation of scenario-generation methods for stochastic programming*. Humboldt-Universität zu Berlin Berlin, 2003.

[24] L. Ntaimo and S. Sen, "The Million-Variable “March” for Stochastic Combinatorial Optimization," *Journal of Global Optimization,* vol. 32, no. 3, pp. 385–400, 2005/07/01 2005, doi: 10.1007/s10898-004-5910-6.

[25] T. A. Santoso, Shabbir; Goetschalckx, Marc; Shapiro, Alexander, "A stochastic programming approach for supply chain network design under uncertainty," *European Journal of Operational Research,* vol. 167(1), pp. 96–115, 2005.

[26] D. Sari and S. M. Ryan, "Statistical reliability of wind power scenarios and stochastic unit commitment cost," (in English), *Energy Syst,* vol. 9, no. 4, pp. 873–898, Nov 2018, doi: 10.1007/s12667-017-0255-7.

[27] K. Gnanendran, K. Iacocca, and R. Lawrence, "A Practical Framework for Aggregate Production Planning," (in English), *Oper Supply Chain Ma,* vol. 19, no. 1, pp. 1–12, 2026. [Online]. Available: <Go to ISI>://WOS:001654872700001.

[28] N. H. Elkasrawy, N. M. Galal, and A. F. Abdelmoneim, "Aggregate Production Planning with Fuzzy Demand and Penalty Costs," (in English), *2020 9th International Conference on Industrial Technology and Management (Icitm 2020),* pp. 112–116, 2020, doi: 10.1109/icitm48982.2020.9080403.

[29] R. J. Hyndman, A. B. Koehler, R. D. Snyder, and S. Grose, "A state space framework for automatic forecasting using exponential smoothing methods," (in English), *Int J Forecasting,* vol. 18, no. 3, pp. 439–454, Jul–Sep 2002, doi: Pii S0169-2070(01)00110-8 Doi 10.1016/S0169-2070(01)00110-8.

[30] K. Høyland and S. Wallace, "Generating Scenario Trees for Multistage Decision Problems," *Management Science,* vol. 47, pp. 295–307, 02/01 2001, doi: 10.1287/mnsc.47.2.295.9834.

[31] S. Mehrotra and D. v. Papp, "Generating Moment Matching Scenarios Using Optimization Techniques," *SIAM Journal on Optimization,* vol. 23, no. 2, pp. 963–999, 2013, doi: 10.1137/110858082.

[32] N. Mortezaei, N. Zulkifli, and M. Nilash, "Trade-off Analysis for Multi-Objective Aggregate Production Planning," *Journal of Soft Computing and Decision Support Systems,* vol. 2, 04/30 2015.

[33] J. D. Mendoza, J. Mula, and F. Campuzano-Bolarin, "Using systems dynamics to evaluate the tradeoff among supply chain aggregate production planning policies," *International Journal of Operations & Production Management,* vol. 34, no. 8, pp. 1055–1079, 2014, doi: 10.1108/ijopm-06-2012-0238.

[34] D. Alem, E. Curcio, P. Amorim, and B. Almada-Lobo, "A computational study of the general lot-sizing and scheduling model under demand uncertainty via robust and stochastic approaches," (in English), *Comput Oper Res,* vol. 90, pp. 125–141, Feb 2018, doi: 10.1016/j.cor.2017.09.005.

[35] S. D. Thompson, D. T. Watanabe, and W. J. Davis, "A comparative study of aggregate production planning strategies under conditions of uncertainty and cyclic product demands," *Int J Prod Res,* vol. 31, no. 8, pp. 1957–1979, 1993/08/01 1993, doi: 10.1080/00207549308956834.

[36] D. G. Gioia, E. Fadda, and P. Brandimarte, "Rolling horizon policies for multi-stage stochastic assemble-to-order problems," (in English), *Int J Prod Res,* vol. 62, no. 14, pp. 5108–5126, Jul 17 2024, doi: 10.1080/00207543.2023.2283570.

[37] R. Cavagnini, F. Maggioni, L. Bertazzi, and M. Hewitt, "A two-stage stochastic programming model for bike-sharing systems with rebalancing," (in English), *Euro J Transp Logist,* vol. 13, 2024, doi: ARTN 10014010.1016/j.ejtl.2024.100140.

[38] C. Ordoudis, V. A. Nguyen, D. Kuhn, and P. Pinson, "Energy and reserve dispatch with distributionally robust joint chance constraints," (in English), *Oper Res Lett,* vol. 49, no. 3, pp. 291–299, May 2021, doi: 10.1016/j.orl.2021.01.012.