Advances in Computer, Signals and Systems (2025)
Clausius Scientific Press, Canada
DOI: 10.23977/acss.2025.090110
ISSN 2371-8838 Vol. 9 Num. 1

# *A Study on Production Decision Making Problem Based on Multi-Stage Stochastic Dynamic Programming*

**Xinyue Tong[1], Jiayi Deng[2], Liang Yan[1], Zhengzheng Ning[1], Jiayi Zhou[1,a,*], Zhaoyang Wang[1]**

[1]*School of Information Engineering, Xi'an Mingde Institute of Technology, Xi'an, China*
[2]*School of Economics and Management, Xi'an Mingde Institute of Technology, Xi'an, China*
[a]*15079187699@163.com*
[*]*Corresponding author*

***Keywords:*** Multi-stage Stochastic Dynamic Programming, Production Decision Making, Bayesian Sequential Probability Ratio Test, Uncertainty, Optimisation Models

***Abstract:*** This study aims to explore the production decision-making problem based on multi-stage stochastic dynamic programming to cope with the many uncertainties faced in modern production management. Firstly, the Bayesian sequential probability ratio test model is built to solve the problem of sampling and testing when purchasing spare parts, which effectively reduces the testing cost and improves the reliability of decision-making. Then, a multi-stage stochastic dynamic planning decision-making model is constructed, which integrally considers multiple stages and various cost factors in the production process to maximise the profit of the enterprise. The results show that the model can effectively deal with the stochastic demand and uncertainty in the production process and provide an optimal production decision-making solution for the enterprise. However, the solving efficiency of the model and its ability to handle large-scale data still need to be improved. Future research will be devoted to optimising the algorithm and expanding the application scope of the model to better adapt to the complex and changing production environment.

## 1. Introduction

In modern production management, production decisions are faced with many uncertainties, such as demand fluctuations, production capacity constraints, and changes in the market environment. These uncertainties make it difficult for traditional static optimisation methods to effectively cope with the dynamic changes in the production process [9], and therefore, optimisation methods based on dynamic planning have become an important tool to cope with such problems. Dynamic planning is able to deal with the time dimension and stochasticity in the decision-making process by means of staged decision-making, thus helping enterprises to make optimal production decisions in uncertain environments.

Multi-stage stochastic dynamic programming (MSDP) further considers the interactions of different time phases in the production process, and can effectively cope with stochastic demand and uncertainty in the production process. In recent years, MSDP has been widely used in the fields of production scheduling and inventory management [8], showing its potential in complex production environments. However, existing studies still face challenges such as solution efficiency and model complexity, therefore, this study aims to explore the production decision-making problem based on multi-stage stochastic dynamic programming and propose optimisation schemes to improve the accuracy of decision-making and implementation.

## 2. Literature Review

Since the 90's, the technical barriers between enterprises are getting lower and lower, and the position of cost in competition is getting more and more important. With the high speed development of information technology and the application of advanced manufacturing methods, the research, application and development of cost variance control methods have also shown a diversity and cross-trend [1]. The problem of cost variance control is also manifested in the application of multidisciplinary and multi-methods, such as fuzzy set theory [2], DSS technology [3] and neural network models [4], and a variety of intelligent methods and statistical variance control methods are combined to jointly carry out analysis and decision-making. The commonality of these methods is to seek the approximate optimal solution or satisfactory solution to meet the actual needs of the optimisation problem of complex manufacturing systems, rather than the exact optimal solution, which is a solution to the optimisation problem of complex manufacturing systems with application prospects [5], but these intelligent methods also have weaknesses such as slow convergence, easy to fall into the local optimal solution and difficult to handle the complex constraints, which require further in-depth research.

Since the twenty-first century, science and technology, especially information technology, have undergone rapid development, the traditional product structure has changed dramatically, and a high degree of automation of production based on high technology and a variety of advanced manufacturing methods oriented to customer needs have been popularised [6,7]. Difficulties in applying statistical methods to real cost environments have been resolved gradually through the use of advanced computer information technology by means of system integration, intelligence and dynamics [7]. Therefore, the focus of this thesis is on the fact that the statistical environment for cost variance has changed today and many old methods have taken on new relevance in the new environment [10].

This paper combines the Bayesian theory of this type of statistical method with the dynamic programming method and uses MATLAB software to assist in calculations based on the actual cost environment for process analysis and decision making of cost variance data, and empirical analysis shows the feasibility and effectiveness of this method.

## 3. Sampling and Testing Issues

When enterprises purchase spare parts, in order to ensure product quality, they need to judge whether the defective rate of spare parts exceeds the nominal value promised by suppliers through sampling and testing. According to the requirements of enterprises, it is necessary to design a sampling and testing programme with as few tests as possible to determine whether the defective rate exceeds the standard or passes the test at 95% and 90% confidence level respectively.

Enterprises need to ensure that the spare parts they purchase meet quality standards to ensure that the finished product is acceptable. The two main challenges faced by companies in the inspection process are as follows:

Uncertainty of the defective rate: the supplier claims that the defective rate of spare parts does not exceed the nominal value of 10%, but the actual defective rate is unknown, and the enterprise hopes to determine whether the defective rate is in accordance with the standard through sampling and testing.

Limitations on testing costs: Companies want to make reliable decisions with as few tests as possible and avoid unnecessary spending on testing costs.

In order to effectively solve the problem of determining the defective rate of enterprises in sampling and testing, the Bayesian Sequential Probability Ratio Test (SPRT) model has been developed. Bayesian inference provides a dynamic estimate of the defective rate, while the sequential probability ratio test (SPRT) utilises this estimate during the inspection process to make real-time acceptance or rejection decisions. This combination reduces unnecessary inspection samples. Enterprises can adjust the prior distribution of Bayesian inference as well as the error probability threshold of SPRT according to the needs of actual production, so as to cope with different production and inspection needs. The testing steps are as follows:

## 3.1 Assumptions

During the sampling process, the firm needs to make a dynamic decision based on the test results: whether to continue sampling, accept the batch of spare parts, or reject the batch of spare parts. Let p be the actual defective rate, which is claimed by the supplier $p \le p_0 = 10\%$, so the following two assumptions are established:

$$\begin{cases} H_0: p \le p_0, \text{Match Condition} \\ H_1: p > p_0, \text{Mismatch Condition} \end{cases} \tag{1}$$

where p0 is the nominal value of the supplier's claimed defective rate. H0 indicates that the supplier's claimed defective rate does not exceed the nominal value; H1 indicates that the defective rate provided by the supplier exceeds the nominal value.

Next, the sampling data is used to determine whether or not to accept the batch of spare parts. In case of uncertainty about the defective rate, the defective rate estimate is first updated by Bayesian inference and then combined with SPRT for decision making.

## 3.2 Bayesian Inference Modelling

In the process of using Bayesian inference, it can be assumed that the substandard rate of spare parts p obeys the Beta distribution, as the prior information on the unknown substandard rate [1].The Beta distribution is a commonly used distribution for dealing with proportional data such as substandard rates, and the conjugacy with the binomial distribution makes Bayesian updating calculations easier. The prior distribution is assumed to be:

$$p \sim B(\alpha_0, \beta_0) \tag{2}$$

$\alpha_0$ and $\beta_0$ denote firms' beliefs about low and high defect rates, respectively.

As the sample testing proceeds, the number of substandard products in the n samples is observed to be X. According to the update rule of Bayesian inference, the posterior distribution of the substandard rate p is:

$$p|X \sim B(\alpha_0 + X, \beta_0 + n - X) \tag{3}$$

We consider a number of different defective rate scenarios, covering situations ranging from below to above nominal values, in order to fully evaluate the system's strategy under different quality batches:

Low Defective Rate ($p < p_0$):

(1) Actual reject rate: Setting p=0.08, p=0.06 and p=0.04 is used to simulate a scenario where the actual reject rate is less than 10% of the nominal value.

(2) Prior distribution: for the low defective rate scenario, the more conservative Bayesian prior distribution is used. Setting $\alpha_0=1$ and $\beta_0=9$ indicates that companies have a strong belief in lower defective rates and tend to receive batches with defective rates below 10 per cent.

High Defective Rate ( $p > p_0$ ):

(1) Actual reject rate: Setting p=0.12, p=0.14 and p=0.16 is used to simulate batches with reject rates exceeding 10 per cent.

(2) Prior distribution: for the high defective rate scenario, the use of a stricter prior distribution $\alpha_0 = 3$ and $\beta_0 = 7$ indicates that companies are cautious about high defective rates and are more inclined to reject high defective rate batches.

### 3.3 Sequential Probability Ratio Test (SPRT)

In order to dynamically judge whether to continue sampling or make acceptance/rejection decisions, the sequential probability ratio test is introduced. SPRT calculates the log-likelihood ratio [2] (LLR) after each sampling to judge the acceptance or rejection of spare parts and makes judgements based on the LLR values with set boundaries to ensure that the conclusion is reached in fewer number of tests, with the following formula:

$$LLR = \log\left(\frac{P(X|H_1)}{P(X|H_0)}\right) \tag{4}$$

Where, $P(X|H_0)$ and $P(X|H_1)$ denote the probability of the number X of inferior products under $H_0$ and $H_1$ respectively. Based on the covariance of Beta and Binomial distributions in the Bayesian inference model, the probability can be expressed as:

$$\begin{cases} P(X \mid H_0) = \binom{n}{X} p_0^x (1-p_0)^{n-x} \\ P(X \mid H_1) = \binom{n}{X} p_1^x (1-p_1)^{n-x} \end{cases} \tag{5}$$

Where n is the number of samples and p1 is an assumed value set greater than the maximum permissible value of the product defect rate. Based on the calculation of the log-likelihood ratio, the rules for companies to make acceptance or rejection decisions are as follows:

If LLR>logA, the defective rate is considered to be significantly higher than p0 and the batch of spare parts is rejected;

If LLR < logB, the defective rate is considered to be no more than p0 and the batch of spare parts is accepted;

If B < LLR < A, continue sampling until a definitive judgement is made.

Where A and B are the upper and lower boundaries for decision making, the boundary values will be used to determine whether the log-likelihood ratio falls in the interval of acceptance, rejection, or continued detection to help companies make real-time decisions. The boundary values A and B are related to the probability of error $\alpha$ and $\beta$ of detection with the following equations:

$$A = \frac{1-\beta}{\alpha}, B = \frac{\beta}{1-\alpha} \tag{6}$$

Where $\alpha$ and $\beta$ denote the tolerance of the enterprise incorrectly rejecting conforming batches, and the tolerance of the enterprise incorrectly accepting non-conforming batches, respectively. It is necessary to further dynamically adjust the parameters of the prior distribution to respond more flexibly to changes in the rate of defective products, choose more stringent rejection criteria, and further lower the thresholds, with the adjustment of $\alpha$ and $\beta$ , and repeat the above steps for the simulation experiments.

With a low defect rate (acceptance) $\alpha = 0.20, \beta = 0.25$ at 95% confidence, the system is more likely to accept qualified batches;

High defect rate (rejections) at 90% confidence $\alpha = 0.005, \beta = 0.02$ , the system rejects failed batches more quickly.

Setting the number of parametric simulations to 1000 and the maximum sample size to 100, the above model was solved using Matlab and the following results were obtained:

In scenario 1, with a defective rate of >10 per cent:

Table 1 Proportion of Acceptance at High Substandard Rates with Mean Sample Size

| Actual Defect Rate | Acceptance Rate | Average Sample Size |
|---|---|---|
| 0.12 | 0.00 | 10.00 |
| 0.14 | 0.00 | 9.00 |
| 0.16 | 0.00 | 8.00 |

In Scenario 2, the defective rate is <10 per cent:

Table 2 Proportion of Acceptance at Low Substandard Rates with Mean Sample Size

| Actual Defect Rate | Acceptance Rate | Average Sample Size |
|---|---|---|
| 0.08 | 0.92 | 1.00 |
| 0.06 | 0.93 | 1.00 |
| 0.04 | 0.95 | 1.00 |

Analysing Tables 1 and 2, it can be seen that with an overall sample size of 100, Scenario 1 would require a sample size of between 8 and 10 to reject the spare parts with 95 per cent confidence; Scenario 2 would require a sample size of 1 to accept the spare parts with 90 per cent confidence.

## 4. Sampling and Testing Issues

During the production process, companies need to make decisions about the quality of parts and finished products, including whether to test parts, whether to test finished products, and how to deal with substandard products. These decisions affect the cost, profitability, and quality management of the product.

### 4.1 Data Pre-processing

Firstly, we converted the data into an easy-to-use form and structured it so that it could be directly input into the model for calculation, ensuring that the units were all in dollars per piece. The processing steps are as follows:

The defective rate, purchase unit price, inspection cost, assembly cost, etc. for each situation are converted into matrix or vector form to facilitate multi-scenario simulation.

Standardised treatment of all types of cost items, including dismantling costs, exchange losses, procurement costs of spare parts, testing costs, assembly costs in each case, to ensure uniformity of inputs

## 4.2 Multi-stage Stochastic Dynamic Programming (SDP) based Decision-making Modelling

A firm's specific production process contains multiple stages, each with different decisions and states. These decisions are made in multiple stages, and each decision not only affects the cost and revenue of the current stage, but also affects the state and decisions of the subsequent stages, and it is necessary to consider possible future states at each stage to make the optimal decisions in production and quality control to maximise the overall profit [3]. The steps to build a multi-stage stochastic dynamic planning model are as follows:

Step1: Define Decision Variables

Decision variables are defined for each stage of the production process of the enterprise (where 1 means detection and 0 means no detection):

$$\begin{cases} x_1 \in \{0,1\}: & \text{Whether to detect the spare parts t,} \\ x_2 \in \{0,1\}: & \text{Whether to detect spare parts 2,} \\ y_f \in \{0,1\}: & \text{Whether the finished product is tested,} \\ z \in \{0,1\}: & \text{Whether to dismantle the unqualified finished product.} \end{cases} \quad (7)$$

Step2: Establish the Objective Function

The profit of a business consists of three components: sales revenue, total production costs and exchange losses. The goal is to maximise profit by adjusting the inspection and dismantling strategy. Profit = Sales Revenue - Total Cost of Production - Exchange Losses. Each component is analysed below:

Sales revenue (R). $$R = 56Q(1 - p_{fi}), i \in (1, 2, \cdots, 6) \quad (8)$$

Where, $p_{fi}$ is the rate of defective finished goods after assembly and Q is the production capacity, which indicates the maximum number of finished goods that a firm can produce per cycle.

Total cost of production ( $C_t$ ).

$$C_t = (C_{1i} + x_1 D_{1i})Q + (C_{2i} + x_2 D_{2i})Q + C_{ai}Q + y_f D_{fi} Q, i \in (1, 2, \cdots, 6) \quad (9)$$

Among them, this part of the cost includes the cost of purchasing spare parts, the cost of testing spare parts, the cost of assembling the finished product, and the cost of testing the finished product: $C_{1i}$ and $C_{2i}$ represent the unit price of purchasing each spare part 1 and spare part 2, respectively; $x_1$ and $x_2$ represent the decision variables of testing spare parts 1 and 2, respectively; $D_{1i}$ and $D_{2i}$ represent the cost of testing spare parts 1 and 2, respectively; $C_{ai}$ is the cost of assembling the finished product; and $y_f$ is the decision variable of testing the finished product.

Loss on exchange ( $L_l$ ) $$L_l = L_i p_{fi} Q, i \in (1, 2, \cdots, 6) \quad (10)$$

Where $L_i$ represents the replacement loss of non-conforming finished products under different scenarios, the replacement loss is proportional to the number of non-conforming products and each non-conforming product will incur an additional replacement cost.

Cost of dismantling ($C_d$) $$C_D = zC_{di}\,p_{fi}Q, i \in (1,2,\cdots,6) \tag{11}$$

where z represents the decision variable of whether or not to dismantle the nonconforming finished product and $C_{di}$ is the dismantling cost of the nonconforming finished product. If disassembly is chosen, additional disassembly costs are incurred.

The above process is analysed for the calculation of the finished product defective rate pf. The defective rate of the assembled finished product is determined by a combination of the defective rates of part 1, part 2 and the defective rate of the finished product itself. The inspection decision affects the calculation of the defective rate. The formula for calculating the defective rate for the finished product is derived as follows:

$$p_F = 1-(1-p_1(1-x_1))(1-p_2(1-x_2))(1-p_{fi}) \tag{12}$$

$C_{1i}$, $C_{2i}$, $D_{1i}$, $D_{2i}$, $C_{ai}$, $C_{di}$, $L_i$, $p_{fi}$ in Table 1 yield different values depending on the situation. Solving the above equation by association yields maximised profit with an objective function:

$$\max P_a = R - C_t - L_l - C_D \tag{13}$$

Step3: Finding Constraints

The constraints on the model are as follows:

$$\begin{cases} x_1 \in \{0,1\}, x_2 \in \{0,1\}, y_f \in \{0,1\}, z \in \{0,1\} \\ Q = \text{Production capacity of enterprises} \end{cases} \tag{14}$$

Step4: Introducing Riskiness (additional conditions)

On the basis of the above, firms are not only concerned with maximising profits but also consider potential risks when making inspection and dismantling decisions, and this risk awareness or decision-making preference can be described by the introduction of a risk control model or the psychological expectations of the decision maker. The above model is adapted as follows:

Risk modelling of transfer losses

Assuming that firms are somewhat risk averse to fluctuations in defective rates or swapping losses, firms may be more concerned about swapping losses from high defective rates, so the risk can be defined and weighted as swapping losses:

$$L_T = (1+\lambda)L_i\,p_{fi}Q, i \in (1,2,\cdots,6) \tag{15}$$

The adjusted objective function is:

$$\max P_a = R - C_t - L_T - C_D \tag{16}$$

Firms will not only consider normal swapping losses, but will also amplify them based on the level of risk aversion $\lambda$. A high level of $\lambda$ indicates that a company is more sensitive to the potential risk of defective products and will make decisions that favour increased testing or dismantling to reduce potential losses.

Risk modelling of finished product defect rates

Firms may also be very sensitive to the rate of defective products and do not want too many defective products to enter the market. We can introduce a risk-control objective on the defective rate in the model:

$$P_a = R - C_t - \lambda p^2{}_{fi}, i \in (1,2,\cdots,6) \tag{17}$$

Where $p^2{}_{fi}$ represents the square of the defective rate in different cases, this is done to highlight the negative impact on the firm when the defective rate is large, when there is no need to introduce the disassembly part directly in the defective rate squared term. Finding the above constraints, including risk constraints (defective rate constraints, swap loss constraints), in addition to adjusting through the objective function, can also be used to limit the enterprise's decision-making through the introduction of risk constraints, the constraints are as follows:

$$\begin{cases} p_{fi} \le p_{\max} \\ L_l \le L_{\max} \end{cases} \tag{18}$$

In this case, the firm sets a maximum acceptable defective rate $p_{\max}$ and a threshold $L_{\max}$ that it wishes to ensure that swap losses do not exceed. With this constraint, the firm is not only able to avoid excessive defective rates, even though the optimal model may result in a profit-maximising solution at higher defective rates, but also limits the losses at high defective rates and avoids excessive switching costs.

On the basis of the above, considering the important condition of cyclicity, in each round of cyclic production, the firm adjusts its strategy for the next round based on the decisions and results of the previous round, in order to expect an optimal overall profit performance. The firm's goal is to ultimately maximise the total expected profit for the entire production cycle by adjusting the decisions in each round in multiple cycles of production and updating the maximum profit objective function. It can be expressed as:

$$P_t = \sum_{t=1}^{T} P_a^t, (t = 1,2,\cdots,T) \tag{19}$$

$$\max P_a^t = 56Q(1 - p^t{}_{fi}) - C^t{}_t - (1+\lambda_1) L_i p^t{}_{fi} Q^t - \lambda_2 (p^t{}_{fi})^2, i \in (1,2,\cdots,6) \tag{20}$$

Where $P_a^t$ represents the adjusted profit in round t. By the same description above the cyclic conditions are introduced: $x^t{}_1$, $x^t{}_2$, $y^t{}_f$, $z^t$, which represent the decision variables at each stage in round t. None of the conditions change. Next, the defective rate and total cost will change, which are $p^t{}_{fi}$ and $C^t{}_t$, respectively.

In each round of production, the firm adjusts the decision variables to maximise overall profitability based on the previous round's production performance and current market conditions. The entire optimisation process can be carried out in the following steps:

Set the initial decision variables, defective rate, and set the number of production according to Table 1;

Calculate the total cost, defective rate, sales revenue, swap loss and adjusted profit for each round.

Testing and dismantling decisions are adjusted based on adjusted profit and risk control factors.

Record the profit and defective rate for each round as a reference for the next round of production.

Terminate the loop when profits converge or the maximum number of loops is reached.

By setting up a multi-stage stochastic dynamic programming and discussing the values of these decision variables while taking into account all the scenarios that arise, the enterprise can decide whether or not to test part 1, part 2, the finished product, and whether or not to dismantle the non-conforming finished product under each stage. These decisions affect the calculation of all types of costs in the production process, which in turn affects the final profit of the firm.

### 4.2 Multi-stage Stochastic Dynamic Programming (SDP) based Decision-making Modelling

Based on the model developed above, the initial decision variables, risk aversion coefficients, cyclic conditions, maximum number of cycles, and tolerances need to be set first, and the results are as follows:

$$\begin{cases} \text{Production capacity Q} = 1000 \\ \text{Risk aversion coefficient } \lambda_1 (\text{enterprise sensitivity to switching losses}) = 1.5 \\ \text{Risk aversion coefficient } \lambda_2 (\text{sensitivity of enterprises to fluctuation}) = 1.0 \\ \text{Maximum number of cycles max\_loops} = 50 \\ \text{Tolerance} = 0.01 \end{cases} \quad (21)$$

Cyclic production process steps (same logic as Figure 1):

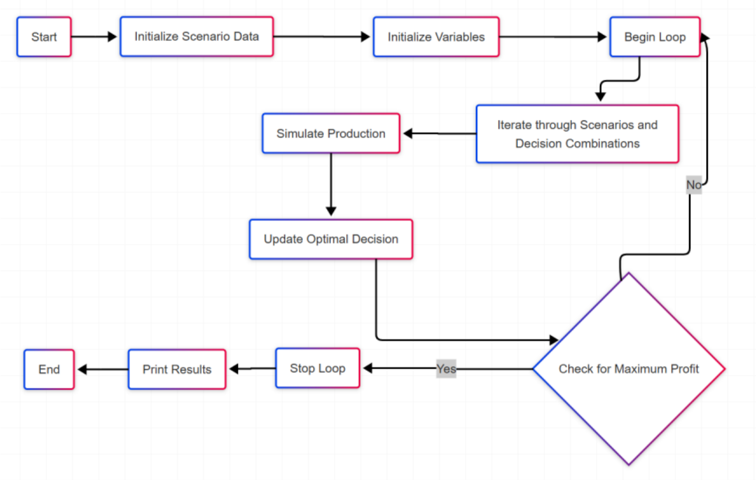


Figure 1. Cyclic production process steps

Step 1: Initialise the parameters including production capacity Q, swap loss factor $\lambda_1$, reject rate factor $\lambda_2$ and initial decision history;

Step 2: Enter the loop and iterate through the production rounds

Step 3: Iterate through the decision combinations, detecting part 1, part 2, and the finished product, and deciding whether or not to disassemble the nonconforming product.

Step 4: Calculate the profit, calculate the profit for each round based on each decision and update the optimal decision.

Step 5: Check the convergence condition and terminate the loop if the change in profit is less than the set threshold.

Step 6: Output results, output the optimal decision and final profit.

Combined with the data in Table 1, the above model was solved using Matlab to derive the decision-making options and the corresponding bases in different cases, and the results are as shown in Table 3:

Table 3 Table of Decision-making Options and Corresponding bases in Different Scenarios

| Gauge | Scenario 1 | Scenario 2 | Scenario 3 | Scenario 4 | Scenario 5 | Situation 6 |
|---|---|---|---|---|---|---|
| Decision-making Programme | [0,0,1,1] | [0,0,1,1] | [0,0,1,1] | [1,1,1,1] | [0,1,1,1] | [0,1,0,0] |
| Margins | 23075.90 | 21535.20 | 20799.50 | 20900.00 | 23385.00 | 23702.11 |
| Defective Rate | 2.71 per cent | 4.88 per cent | 2.71 per cent | 2.00 per cent | 1.90 per cent | 1.43 per cent |
| Total Costs | 31000.00 | 31000.00 | 31650.40 | 32480.00 | 31076.00 | 31142.62 |
| Sales Revenue | 54482.40 | 53267.20 | 54482.40 | 54880.00 | 54936.00 | 55201.30 |
| Replacement and Recovery Costs | 162.60 | 292.80 | 813.00 | 600.00 | 190.00 | 142.63 |

Note: Total costs, sales revenue, replacement and recovery costs, and profit are shown in dollars in the table.

## 5. Summary

In this study, the production decision-making problem is discussed in depth through the establishment of a multi-stage stochastic dynamic planning model, which provides an optimal decision-making solution for enterprises in complex and changing production environments. The results show that the model can effectively deal with the uncertainties in the production process and improve the accuracy and implementation of decision-making. However, there are still some limitations in the current study, such as the solving efficiency of the model needs to be further improved, and there may be some computational bottlenecks when dealing with large-scale production data. Future research will be devoted to optimising the model's algorithm to enhance its applicability and efficiency in large-scale production scenarios. In addition, more complex factors in actual production, such as supply chain fluctuations and equipment failures, will be explored to be included in the model to further expand the application scope of the model and provide more comprehensive and accurate production decision support for enterprises.

## References


*[1] Zou B, Peng J, Li S, et al. Comparative study of the dynamic programming-based and rule-based operation strategies for grid-connected PV-battery systems of office buildings[J]. Applied energy, 2022, 305: 117875.*

*[2] Wu M, Xu J, Zeng L, et al. Two-stage robust optimisation model for park integrated energy system based on dynamic programming[J]. Applied Energy, 2022, 308: 118249.*

*[3] Kastius A, Schlosser R. Dynamic pricing under competition using reinforcement learning[J]. Journal of Revenue and Pricing Management, 2022, 21(1): 50-63.*

*[4] Li M, Mizuno S. Dynamic pricing and inventory management of a dual-channel supply chain under different power structures [J]. European Journal of Operational Research, 2022, 303(1): 273-285.*

*[5] Chen B, Simchi-Levi D, Wang Y, et al. Dynamic pricing and inventory control with fixed ordering cost and incomplete demand information [J]. Management Science, 2022, 68(8): 5684-5703.*

*[6] DemirovićE, Lukina A, Hebrard E, et al. Murtree: optimal decision trees via dynamic programming and search[J]. Journal of Machine Learning Research, 2022, 23(26): 1-47.*

*[7] Liu H, Wang Y, Lee L H, et al. An approximate dynamic programming approach for production-delivery scheduling under non- stationary demand [J]. Naval Research Logistics (NRL), 2022, 69(4): 511-528.*

*[8] Chouaf W, Abbou A, Agga A. Optimal energy management for a connected microgrid using dynamic programming method[C]//2022 8th International Conference on Optimisation and Applications (ICOA). IEEE, 2022: 1-6.*

*[9] Wang P, Cao J, Liu X, et al. Research on Enterprise Order and Transshipment Strategy Based on Factor Analysis and Dynamic Programming[J]. Academic Journal of Mathematical Sciences, 2022, 3(2): 13-21.*

*[10] Fennich M E, Fomeni F D, Coelho L C. A novel dynamic programming heuristic for the quadratic knapsack problem [J]. European Journal of Operational Research, 2024, 319(1): 102-120.*