Summary of China Case Studies in WP1

A mini-hub refers to a warehousing and distribution facility in the logistics chain that connects central warehouses with end customers and supports last-mile delivery. Community mini-hubs are typically located within 3 kilometers of consumers to shorten delivery distances and time, enabling minute-level instant delivery. As an asset-heavy model, mini-hubs are highly vulnerable to fluctuations in market demand and prices. Thus, the scientific and rational layout of mini-hubs and the optimization of their capacity are critically important.WP1: Locations of Fixed and Mobile Mini-Hubs includes three tasks: Task 1.1: Determination of Candidate Mini-Hub Locations, Task 1.2: Mini-Hub Location Planning, Task 1.3: Stage-wise Deployment of Mini-Hubs.

Task 1.1 Candidate mini-hub location determination

In this study, we propose a two-stage framework for selecting mini-hub candidate locations, combining Improved Fuzzy C-means Clustering and the Hesitant Fuzzy TOPSIS method. Each point is treated as the basic unit, and points with similar characteristics are grouped into clusters. We then apply Hesitant Fuzzy TOPSIS to each cluster to identify candidate mini-hub locations.

Based on theoretical analysis and questionnaire surveys, we define multiple criteria for mini-hub development, including coordinates, construction cost, population, POI information, road facilities, and distance to residential areas.

The selected Chinese case is Huangdu Town, Jiading District, Shanghai. In 2023, it had a total population of 116,409, covering 28 communities, with diverse land-use types and uneven population distribution. The administrative map of Huangdu Town is shown in Figure 1(a).

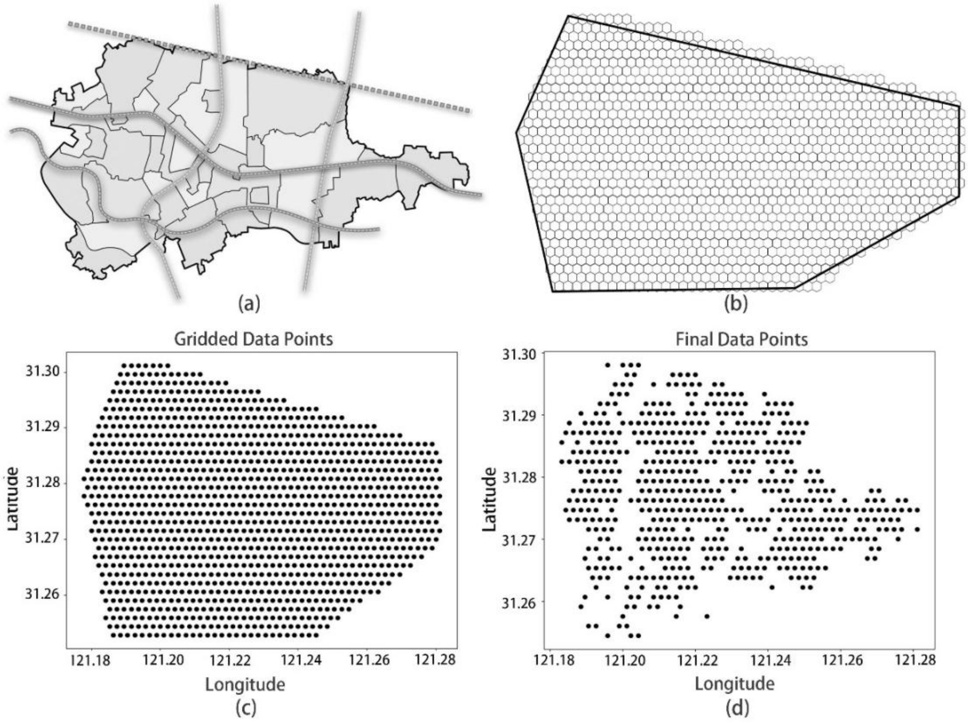


Figure 1. Empirical case data

As shown in Figure 1(b), we divide the entire study area using a hexagonal grid. We assume that conditions relevant to mini-hub development within a 100-meter radius are approximately homogeneous. We then extract the center of each grid to form sample points, as illustrated in Figure 1(c).

Figures 1(c) and 1(d) present the spatial distribution of sample points, with longitude on the X-axis and latitude on the Y-axis. Some points are excluded due to water bodies, mountainous terrain, or regulatory restrictions. The remaining valid points are shown in Figure 1(d), which form the final candidate locations.

Figure 2 shows the spatial distribution of sample points, where points marked with "×" represent candidate locations. The distribution is reasonable and covers the whole region, ensuring service quality and supporting subsequent refined location planning research.

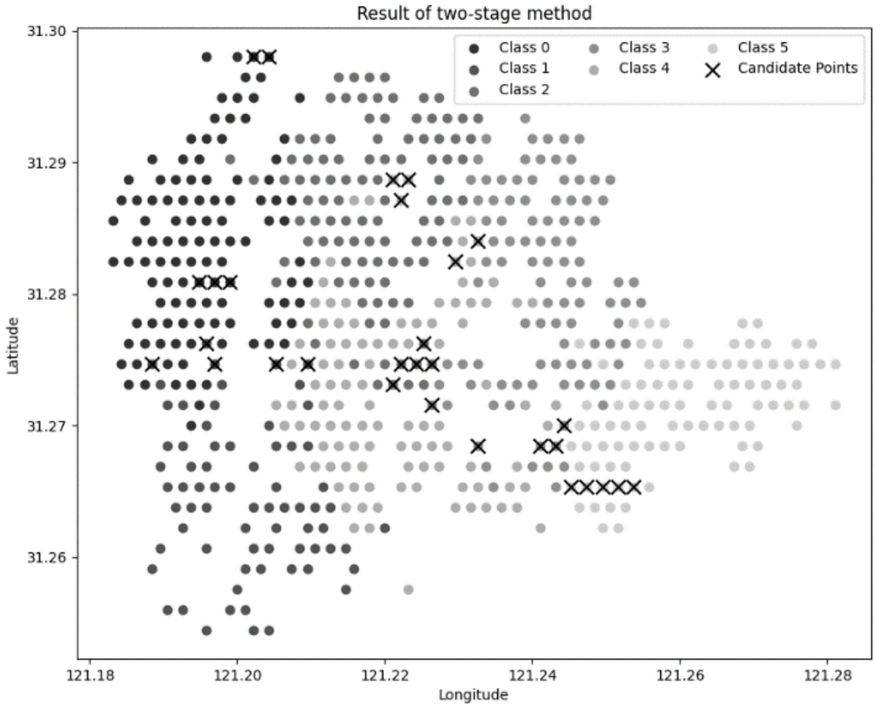


Figure 2. Result of two-stage method

Task 1.2 Mini-hub location planning

This study extends Task 1.1 by introducing mobile mini-hub delivery, which supports low-cost, long-distance bulk delivery while considering stochastic demand fluctuations. The overall problem is formulated as a two-stage stochastic programming model.In the first stage, we model mini-hub location planning. In the second stage, we optimize rider allocation, self-pickup and delivery assignments for fixed hubs, and routing for mobile mini-hubs.

Given the large data size and multiple delivery modes, the full two-stage model is highly complex and difficult to solve directly. We therefore decompose the problem. The optimal delivery plan for mobile mini-hubs is treated as a subproblem, which captures the relationship between mobile mini-hub delivery profit and total distribution cost. The solution is based on the column generation algorithm combined with the bidirectional labeling method.

We also reformulate the first-stage master problem: a genetic algorithm generates and filters multiple combinations of candidate locations, which are evaluated with a fast second-stage algorithm to identify the optimal location planning scheme.

The study area remains Huangdu Town. Final outputs include specific mini-hub locations and daily delivery routes for mobile mini-hubs. The location planning scheme is illustrated in Figure 3, where points marked with red “×” represent the optimal locations.

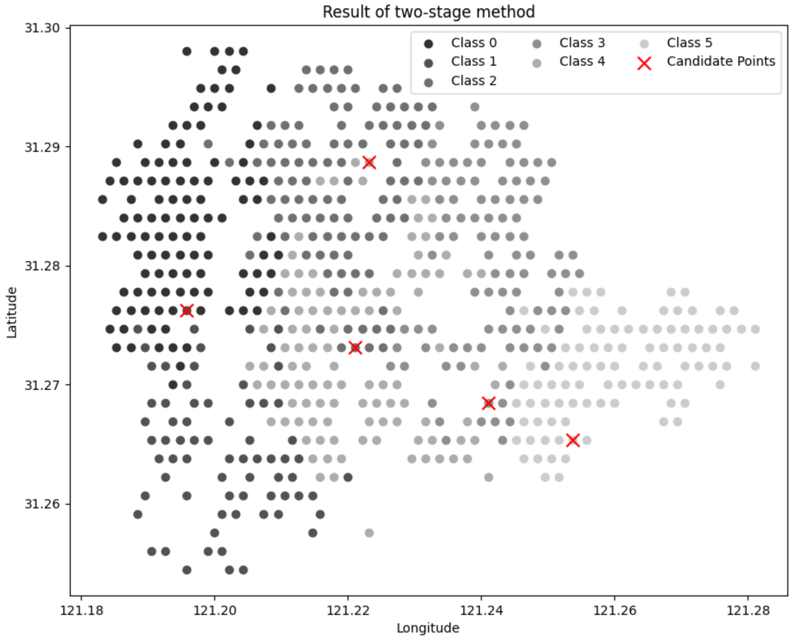


Figure 3. Result of location plan

Task 1.3 Stage-wise deployment of mini-hubs

We model the problem as a Markov Decision Process (MDP). The location of the regional central warehouse, which supplies mini-hubs, is fixed throughout the process.

In the single-stage model, mobile mini-hubs are not considered. The objective is to minimize costs from mini-hubs to each demand zone. The output is the optimal set of mini-hub locations from the candidate locations and the zones they serve.

In multi-stage planning, the regional central warehouse remains fixed, and some customers can pick up orders at their nearest mini-hub.

Again, Huangdu Town is used as the study area. For the single-stage problem, by appropriately relaxing delivery volume constraints, the Gurobi solver achieves satisfactory accuracy within a short time, meeting solution quality requirements. The feasibility of the single-stage model has been verified, and we are currently extending the framework to multi-stage modeling and solution.

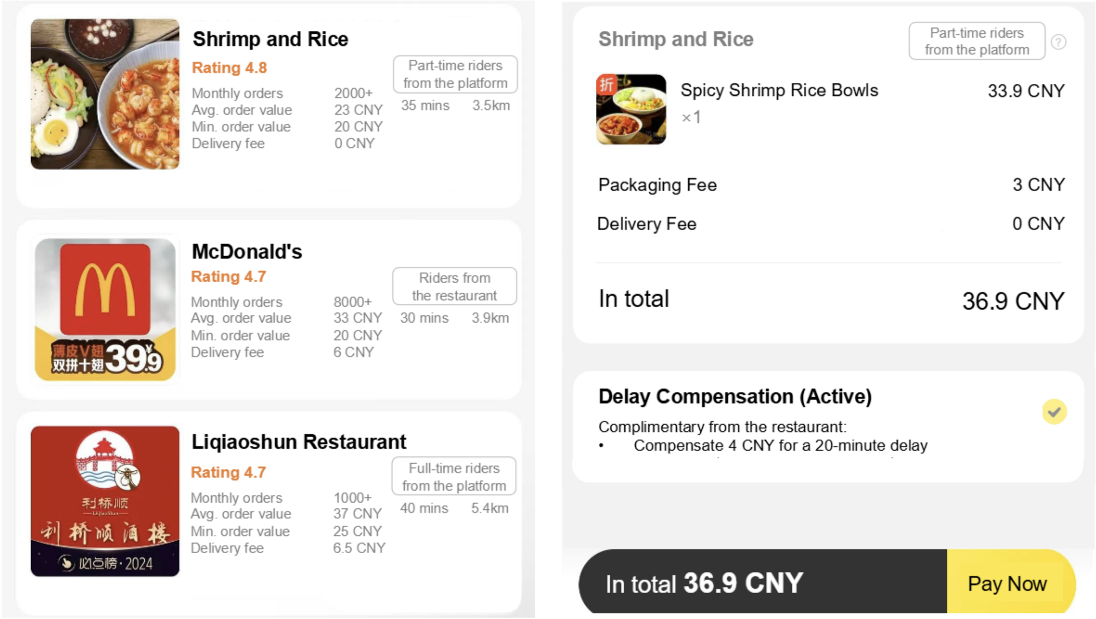
Summary of China Case Studies in WP3

Task 3.1: Dynamic Assortment Optimization for Online Food Delivery

Context and Objectives

In recent years, it has become the norm for Chinese restaurants to offer Online Food Delivery (OFD) services alongside their traditional dine-in options to expand their customer base and improve profitability. The scale of this market is massive; by the end of 2023, over 3 million restaurants in China had served 545 million customers through OFD. Major Chinese platforms, such as Meituan and Ele.me, serve as the primary intermediaries, coordinating vast fleets of riders to fulfill orders.

Despite efforts to ensure timely deliveries, delays still occur due to traffic congestion, extreme weather, and insufficient delivery resources, leading to customer dissatisfaction. To address this, Chinese platforms and restaurants have collaboratively introduced Delay Compensation (DC), a complimentary service where customers are financially compensated if their delivery is delayed beyond a specific threshold. Figure 1 presents representative user interfaces from a major Chinese OFD platform when DC is offered.



Shrimp and Rice
Rating 4.8
Monthly orders 2000+
Avg. order value 23 CNY
Min. order value 20 CNY
Delivery fee 0 CNY
Part-time riders from the platform
35 mins 3.5km

McDonald's
Rating 4.7
Monthly orders 8000+
Avg. order value 33 CNY
Min. order value 20 CNY
Delivery fee 6 CNY
Riders from the restaurant
30 mins 3.9km

Liqiaoshun Restaurant
Rating 4.7
Monthly orders 1000+
Avg. order value 37 CNY
Min. order value 25 CNY
Delivery fee 6.5 CNY
Full-time riders from the platform
40 mins 5.4km

Shrimp and Rice
Part-time riders from the platform
Spicy Shrimp Rice Bowls ×1 33.9 CNY
Packaging Fee 3 CNY
Delivery Fee 0 CNY
In total 36.9 CNY

Delay Compensation (Active)
Complimentary from the restaurant:
- Compensate 4 CNY for a 20-minute delay

In total **36.9 CNY**

(a) List of restaurants.

(b) Order details.

Figure 1: Representative user interfaces from a major Chinese OFD platform when DC is offered.

While DC alleviates customer concerns, it potentially increases operational costs. Thus, empirical research quantifying heterogeneous customer preferences and their Willingness To Pay (WTP) for DC is critically important for stakeholders in the Chinese OFD market. To facilitate this assessment, our case study addresses this gap by estimating customers' WTP for DC, analyzing its interaction with other key delivery attributes (such as rider type), and investigating the underlying sources of preference heterogeneity.

Survey Design and Modelling Framework

To collect the empirical data, we conducted a stated preference survey administered online via a professional survey firm in China during February and March 2024. The core of the survey was a discrete choice experiment where respondents evaluated hypothetical delivery options. For each delivery option, we designed five distinct attributes, including rider type, delivery time, delivery fee, delay threshold and compensation amount. Table 1 presents a sample choice task.

Table 1: Sample choice task in the survey.

Settings	Please choose your preferred delivery option	
	Option 1	Option 2
Rider type	Part-time riders from the platform	Riders from the restaurant
Delivery time	40 minutes	60 minutes
Delivery fee	6 CNY	4 CNY
Delay compensation	Compensate 4 CNY for a 20-minute delay	Not provided

The survey was structured in three parts. First, a screening phase verified that respondents had ordered OFD within the past seven days to mitigate recall bias, capturing details of their most recent order (e.g., location, payer, and time). Second, respondents completed six choice tasks (five assigned plus one consistency check) assuming the context of their recent order. Finally, socioeconomic data were collected. After strict filtering, the dataset yielded 286 valid Chinese respondents and 1,430 choice observations.

Based on the collected data, we use a two-step modeling framework to empirically evaluate customer preferences for DC and its interaction with other delivery attributes. Figure 2 illustrates this two-step modeling framework. First, we use the hierarchical Bayes approach to calibrate a Mixed Logit model, which accommodates random taste variations and captures the base preference for the compensation service. Second, we extract individual posterior estimates and apply a linear regression model to explicitly map these variations to observable customer characteristics and usage scenarios.

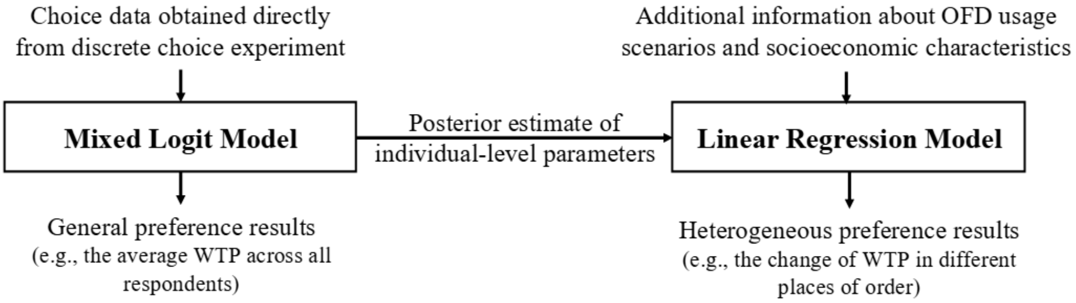


Figure 2: Two-step modeling framework used in this study.

Outcome Analysis

As evaluated by the Mixed Logit model, the results show that under a representative industry scenario (a standard 10-minute delay threshold with platform-affiliated riders, aligning with practices like Meituan's 8-minute delay buffer), the average WTP for DC is 1.20 Chinese Yuan

(CNY). This represents a substantial perceived value, equivalent to roughly 36% of the average delivery fee in China, which is approximately 3.35 CNY. Furthermore, customers prioritize the existence of the DC guarantee and a tight delay threshold over the actual monetary magnitude of the compensation.

We also observe that WTP is highly sensitive to specific delivery attributes, particularly rider type. When orders are fulfilled by restaurant-affiliated riders instead of platform riders, the WTP increases by an average of 0.89 CNY. Customers likely perceive restaurant fleets as less standardized or trackable compared to the highly regulated Meituan or Ele.me fleets, thereby valuing the insurance mechanism more highly.

Additionally, setting the delay threshold too high triggers a detrimental signaling effect. A lenient threshold (e.g., compensating only after a 25-minute delay) signals low service reliability rather than added value, resulting in negative WTP (disutility) among 58.4% of respondents. For detailed summary statistics of WTP under different delay thresholds, please refer to Table 2.

Table 2: Summary statistics of WTP under different delay thresholds with platform-affiliated riders (CNY).

Delay threshold (τ)	Min	25%-Quantile	Median	75%-Quantile	Max	Mean	S.D.
10 min	-1.24	0.77	1.09	1.52	3.80	1.20	0.72
15 min	-1.87	0.28	0.64	1.07	3.47	0.75	0.78
20 min	-2.51	-0.21	0.19	0.66	3.15	0.31	0.83
25 min	-3.14	-0.69	-0.26	0.23	2.82	-0.14	0.88

Using the two-step framework's linear regression, we further trace preference heterogeneity back to specific usage scenarios and socioeconomic traits. When an order is placed at the workplace rather than at a residence, the average WTP is 19.9% higher due to tighter meal schedules. Similarly, WTP is significantly higher for reserved deliveries, larger group orders, and orders paid for personally rather than by a company. Demographically, customers who live in purchased houses (WTP +20.6%), who are married (+17.6%), and who are frequent OFD users ordering more than 6 times a week (+17.0%) exhibit significantly higher valuations for the DC service.

Task 3.2: Dynamic Assortment Optimization for Online Food Delivery

Context and Objectives

In recent years, Online Food Delivery (OFD) has become a core sales channel for Chinese restaurants. Platforms such as Meituan and Ele.me aggregate very large numbers of merchants, meal bundles and dish variants, while customers make decisions under strong time pressure during lunch and dinner peaks. In this environment, restaurants and platforms face a practical assortment problem: they do not have knowledge of the customer preferences.

The decision process in OFD is naturally hierarchical. A customer may first choose among restaurants or cuisine brands, then focus on a meal family such as rice bowls, noodles or burgers, and finally select a concrete item or bundle configuration. This decision process makes the OFD setting a good application for the multi-level nested logit (MLNL) model, which can represent substitution both within a restaurant and across competing merchants.

This case study adapts the dynamic MLNL assortment framework to a Chinese takeout scenario. The managerial objective is to learn customer preferences, and simultaneously refine the dishes highlighted on the app to maximize the expected revenue. To keep the application concrete, the numerical experiments are interpreted as four OFD deployment settings ranging from a regional pilot shelf to a large cross-merchant promotion page.

Case Setting and Modelling Framework

We model the offered OFD catalogue as a three-level MLNL model. At the first level, customers screen among featured merchants or storefronts; at the second level, they compare meal families within a chosen merchant; at the third level, they choose among concrete items such as regular, large or combo versions. In this interpretation, an assortment is the subset of dishes and bundles. Table 1 shows the structure of MLNL model in four cases.

Table 1: Numerical experiment scenarios interpreted as OFD deployment settings

Case	Hierarchy (D1, D2, D3)	Number of items	OFD interpretation
A	(2, 3, 10)	60	2 merchants, each with 3 meal categories, and each category containing 10 specific dish
B	(3, 4, 10)	120	3 merchants, each with 4 meal categories, and each category containing 10 specific dish
C	(3, 4, 5)	60	3 merchants, each with 4 meal categories, and each category containing 5 specific dish
D	(4, 5, 5)	100	4 merchants, each with 5 meal categories, and each category containing 5 specific dish

Outcome Analysis

The first empirical result concerns parameter estimation. Across all four hierarchy structures, the relative L2-norm error of the estimated MLNL parameters decreases steadily as the number of estimation cycles grows, and it trends toward zero over 10^5 cycles. In OFD terms, this means the platform can progressively recover both the base attractiveness of dishes and the substitution strength between related meal variants, even when customers choose among deeply nested restaurant-menu structures.

The more important managerial result is dynamic assortment performance. Against a standard explore-then-exploit baseline that spends a fixed initial phase learning and then commits to one estimated-best assortment, the proposed algorithm delivers consistently lower cumulative regret. At a selling horizon of $T = 10^6$, median regret reductions range from 45.5% to 57.6% across the four cases, while reductions in the maximum regret range from 46.3% to 58.6%. The gains are especially valuable in OFD operations because meal popularity can shift quickly over the day, making a one-shot exploration phase operationally fragile.

Table 2: Performance summary of regret across the four OFD hierarchy cases.

	Baseline		Our Algorithm		Reduction	
Case	Median	Maximum	Median	Maximum	Median	Maximum
A	4334.74	10632.55	2362.47	5261.78	45.5%	50.5%
B	6659.88	17016.15	3397.55	9142.38	49.0%	46.3%
C	7827.85	18266.86	3321.82	7564.19	57.6%	58.6%
D	5200.47	12046.69	2517.57	6101.62	51.6%	49.4%

In summary, the case study shows that the dynamic assortment policy under the MLNL model is useful when the takeout assortment is both broad and structured. In restaurant groups, cloud kitchens or platform campaign pages, customers do not evaluate options as isolated items: they compare merchants, meal families and bundle variants jointly. An online learning policy that respects this hierarchy can therefore support better top-slot allocation, combo recommendation and featured-menu design while producing more stable revenue performance over time.

Task 3.3: Integrated Assortment, Pricing, and Positioning Optimization

Context and Objectives

While optimizing assortments in isolation provides valuable theoretical insights, the reality of Chinese Online Food Delivery (OFD) platforms like Meituan and Ele.me is far more complex. In practice, platforms do not merely decide which dishes to display (Assortment); they simultaneously determine how much to charge for them (Pricing) and where to place them on the screen (Positioning). These three decisions are inherently interdependent.

The "Positioning" decision is particularly critical in the Chinese context due to the "Thumb-scroll" nature of mobile apps. Dishes placed at the top of a category (high visibility) naturally garner more attention, but this effect is modulated by the "Pricing" strategy—discounted items often act as magnets. Conversely, the value of adding a new dish (Assortment) depends on its price point and whether there is a prominent slot available to feature it.

This case study applies the MLNL joint assortment-price-position optimization framework to a Chinese takeout scenario. The managerial objective is to maximize the expected revenue. Based on the conclusions derived from the problem modeling and decomposition, the optimal expected revenue can be expressed as a function of the total price-independent utility. In the following sections, we evaluate the performance of our proposed algorithm in this task by focusing on the computation of this specific quantity.

Case Setting and Modelling Framework

We model the offered OFD catalogue as a three-level MLNL model. In the context of homepage recommendations, the browsing process does not follow an explicit hierarchical path of filtering by store type, then merchant, and finally dishes. Instead, various dishes and meal combos are presented in a unified stream where they share implicit associations. Consequently, the structural parameters of the upper levels are not fixed; only the number of items at the third level remains constant. Therefore, the decision-maker is tasked with determining the optimal assortment and positioning—specifically, selecting which dishes or combos to display and assigning them to specific slots on the page. Table 1 shows the summary of structure parameters and interpretation of four cases.

Table 1: Summary of structure parameters and OFD interpretation.

Case	**Number of positions**	**Number of items**	**OFD interpretation**
A	6	15	15 specific dish to decide for promotion and total 6 available positions
B	8	15	15 specific dish to decide for promotion and total 8 available positions
C	8	20	20 specific dish to decide for promotion and total 8 available positions
D	10	20	20 specific dish to decide for promotion and total 10 available positions

Theoretical Results and Empirical Analysis

From a theoretical perspective, we have established that the problem under study admits a Polynomial-Time Approximation Scheme (PTAS) under mild conditions, providing a solid foundation for the algorithm's feasibility. In the context of this case study, the theoretical relative error bound guaranteed by our algorithm is 0.4. However, a primary objective of this case study is to examine the tightness of this theoretical bound in practical applications. Our empirical results demonstrate that while the theoretical error bound holds mathematically, it is not tight in terms of actual numerical performance. More importantly, compared to the exact algorithm, our approach significantly reduces computational time complexity while ensuring high-quality solutions. The detailed performance comparison is presented in Table 2.

As shown in Table 2, as the problem scale expands from Case A to Case D, the average time consumption of the exact algorithm increases sharply, whereas our algorithm consistently maintains a low computational cost. In terms of time efficiency, our algorithm demonstrates a significant advantage, with the reduction in time consumption ranging from 81.40% to 98.46%. This indicates that the larger the problem scale, the more pronounced the acceleration effect of our algorithm. Regarding solution quality, the actual Mean Absolute Percentage Error (MAPE) is substantially lower than the theoretical error bound of 0.4, with MAPE values across all test cases remaining stable between 0.08 and 0.09. This phenomenon provides strong evidence that our algorithm not only substantially improves computational efficiency but also yields high-quality solutions that are very close to the optimum in practical applications, thereby confirming the looseness of the theoretical error bound and the superior performance of the algorithm in real-world scenarios.

Table 2: Performance summary of time consumption and relative error.

	Average time consumption (/sec)			Relative error
Case	Exact algorithm	Our algorithm	Reduction%	MAPE
A	0.43	0.08	81.40%	0.0808
B	1.76	0.16	90.91%	0.0886
C	11.94	0.43	96.40%	0.0819
D	93.91	1.45	98.46%	0.0812