RAIRO-Oper. Res. 58 (2024) 4307–4340
https://doi.org/10.1051/ro/2024152 **RAIRO Operations Research**
www.rairo-ro.org

THE OPTIMAL DECISION OF SERVICE PROVIDER CONSIDERING EXTRA WAITING AREA VALUE-ADDED SERVICES – POOLED OR DEDICATED?

Zhan Wentao[1], Liu Jida[1], Yuan Xuchuan[2], Huang Xiaotong[1] and Jiang Minghui[1,*]

Abstract. Service providers often provide chargeable extra waiting area value-added services (EWS) to enhance customers' waiting experiences, *e.g.*, in VIP and business lounges at airports, train stations, luxury brand stores, etc., to improve profitability. The impact of EWS on customers' value perception is subject to their heterogeneous sensitivities and also depends on the reference effect due to EWS' observable status. This paper investigates the optimal level of EWS and price decisions for a service provider facing heterogeneous customers, namely the EWS-preferring (E-type) customers and ordinary (O-type) customers, considering the observable status (observable *vs.* unobservable) of EWS using pooled or dedicated capacities in operations through $M/M/1$ queue models. The optimal price and level are solved and the optimal profit is calculated under each scenario. It is found that customers' heterogeneous sensitivities play an important role in the price and intensity decisions. Offering observable EWS to leverage the reference effect in customers' value perception can lead to a higher profit. Furthermore, using pooled capacity in operations with EWS is more profitable when service costs are higher. These results offer significant managerial implications and provide practical guidelines for service providers regarding the intensity of EWS, service price, and whether EWS should be observable to customers through pooled or dedicated capacities.

Mathematics Subject Classification. 90B15.

Received November 20, 2023. Accepted July 18, 2024.

1. Introduction

Value-added services in the waiting area refer to the various additional services provided by service providers to enhance customer utility during the waiting process. For example, at various locations of Haidilao Hotpot, customers are provided with complimentary services such as manicures, shoe shining, and snacks in the waiting area, while some Mystic South-Yunnan Ethnic Cuisine restaurants provide Yunnan-themed performances for customers in the waiting area [33]. These value-added services effectively alleviate negative emotions of customers during the waiting process and enhance their service satisfaction [23,45,47]. Yuan *et al.* [45] conducted the first study on entertainment in waiting areas, which can be seen as a form of value-added service in the waiting area.

Keywords. Extra waiting area value-added services, pooled capacity, dedicated capacity, customer heterogeneity, reference theory.

[1] School of Management, Harbin Institute of Technology, Harbin150001, China.

[2] School of Business, Singapore University of Social Sciences, 599494, Singapore.

*Corresponding author: jiangminghui@hit.edu.cn

They provided the impact function of this value-added service on customers' sensitivity to delays, providing a theoretical basis for this study. In practice, due to limited service resources, not all service providers offer free waiting area value-added services. Instead, these services are offered as additional purchasing options. For example, many airlines have VIP lounges at airports that provide premium waiting environments and exquisite meals to customers who purchase VIP lounge services. Similarly, high-speed rail stations have business waiting rooms that offer better waiting area services to business travelers who pay extra. The world's top luxury stores spare no effort to provide their VIP customers with more valuable added services during their waiting process [29]. These paid waiting area value-added services meet the high needs of some customers for the waiting experience, and this paper refers to them as Extra Waiting area value-added Services (EWS).

However, once service providers only offer EWS to customers who have purchased them (E-type customers), it may lead to a sense of unfairness among ordinary customers. Because when ordinary customers realize that they are being treated differently with inferior service, they perceive unfairness, which further leads to a decrease in their demand [18]. For example, some banks providing seats only for VIP customers have sparked protests from ordinary customers [25]. This sense of unfairness stems from the reference behavior among heterogeneous customers, and the presence of such reference behavior is correlated with the observability of service windows [22]. For instance, when EWS provided to E-type customers cannot be observed by ordinary customers, ordinary customers cannot perceive the value-added services enjoyed by E-type customers in the waiting area, thus losing their reference point and not generating a sense of unfairness based on reference behavior. Similarly, when E-type customers cannot observe the status of ordinary customers in the waiting area, their reference point also disappears, so they will not generate corresponding feelings of superiority and show-off psychology based on reference behavior, leading to a decrease in perceived utility for such customers. Therefore, according to the study by Liu *et al.* [22], we assume that reference behavior among customers only occurs when the provided EWS is observable, and based on this, we explore whether service providers should offer observable EWS.

When there are heterogeneous customers in the market, service providers also need to address the issue of whether to use dedicated queues or pooled queues for service. In practice, some airlines provide VIP lounge services that not only include a premium environment but also have dedicated staff to guide customers to the VIP security checkpoint to reduce the time spent on security checks [34]. This operational approach provides dedicated queues for customers who have purchased EWS, meaning that service providers will allocate dedicated capacity to serve these customers. This helps to enhance the service utility for this category of customers but also requires the service provider to invest more capacity. Conversely, if service providers use pooled capacity to serve both types of customers, it can effectively utilize the available service capacity in the system [37]. It is evident that both using pooled capacity and dedicated capacity have their advantages and disadvantages. Therefore, for service providers operating EWS, how to make a rational choice becomes an important issue.

In response to the research background outlined above, this paper primarily addresses the following questions:

- Should service providers offer observable or unobservable EWS?
- Can providing dedicated queues for customers who purchase EWS lead to higher profits?
- How does the level of heterogeneity among customers affect service provider decisions?
- Does the choice between pooled and dedicated queues vary with the observability of EWS?

To address the above-mentioned questions, this paper constructs a decision model for service providers offering EWS based on customer behavior using an $M/M/1$ queuing model. Firstly, we classify different customers based on their preferences for EWS and provide the effective arrival rates for each customer segment. Then, we study the optimal decisions of service providers in pooled queues when offering observable and unobservable EWS, analyze the impact of customer behavior on service provider profits, and compare the differences in service provider decisions and profits between observable and unobservable EWS. Furthermore, we extend the model to the scenario of dedicated queues and analyze the optimal decisions of service providers when using dedicated capacity for different customer segments. This provides theoretical support for service providers in making service decisions by comparing the profits of service providers using pooled capacity and dedicated capacity.

The main innovations of this paper are as follows: (1) The paper focuses on extra value-added services in waiting areas, expanding the current research field on value-added services in waiting areas. (2) When service providers offer observable EWS, the paper introduces reference theory to analyze the impact of heterogeneous customer behavior on service decisions, addressing the issue of observability of EWS provided by service providers. (3) The paper investigates whether service providers should offer dedicated queues for E-type customers, expanding the application of dedicated capacity and pooled capacity in value-added services in waiting areas. (4) The research finds that the level of heterogeneity in customer reference sensitivity will affect the service providers' choice of EWS observability, while the cost advantage is a key factor determining whether service providers adopt dedicated capacity to serve different customers.

The subsequent sections of this paper are arranged as follows: Section 2 provides a review of the relevant literature. Section 3 constructs the utility function for heterogeneous customers and the decision-making process for service providers based on the base assumptions. Section 4 analyzes the pricing and EWS level decisions for heterogeneous customers when service providers use pooled capacity, and explores the differences in decisions and profits between unobservable and observable scenarios. Section 5 examines the optimal pricing and EWS level when using dedicated capacity, and compares the optimal decisions and profits under different EWS observabilities. Section 6 contrasts the scenarios of using pooled capacity and dedicated capacity. Section 7 concludes the paper. Appendix A provides relevant auxiliary conclusions, while Appendix B offers the derivation process of the propositions and corollaries in this paper.

2. Literature review

This study is primarily related to research on value-added services in waiting areas, reference theory, queue pooling, and delay announcements.

In practice, value-added services in waiting areas mainly include providing high-quality waiting environments and various services. In the research on service environments, Kotler [17] was the first to define "atmospherics" as the service environment that enhances customers' purchase intentions. Subsequently, scholars have studied the changes in customer emotions, dwell time, and payment expectations based on the physical attributes of the service environment [6, 19]. The service environment not only includes the physical aspect but also the social aspect, such as the service level and attitude of service personnel, which have a significant impact on customer consumption behavior [4]. In the related research on waiting area services, Yuan *et al.* [45] first introduced entertainment services in waiting areas into business operations decision-making. They believe that entertainment in waiting areas can reduce customers' unit waiting costs and determine the optimal decisions in cooperative and competitive scenarios. Li *et al.* [20] further expanded on this by analyzing the impact of entertainment in waiting areas on service capacity and providing optimal service capacity. In addition, Sun *et al.* [36] studied a queuing system with an observable queue, where servers provide entertainment services to waiting customers. Based on this, they also developed a game theory model to study the equilibrium behavior of customers. Their analysis indicates that entertainment in the waiting area is attractive only when the market size is moderate and the optimal capacity for entertainment is a unimodal function of the market size.

Currently, research on value-added services in waiting areas often assumes that service providers offer them for free, with limited research on extra value-added services in waiting areas. Therefore, this paper mainly focuses on the impact of extra value-added services in waiting areas on service providers' operational decisions, expanding the current research scope on value-added services in waiting areas. Meanwhile, it provides theoretical support for the operation of extra value-added services in practice, such as VIP lounges in airlines and business waiting rooms in train stations.

Customer consumption behavior exhibits reference dependence characteristics. Customers make consumption decisions not only based on the intrinsic value of products or services but also by comparing the value of products and services with their psychological reference value to determine their final consumption behavior [8, 44, 49]. Tversky [39] were the first to study customers' reference dependence behavior and proposed prospect theory. They further conducted in-depth research on reference-dependent preference models [40]. Reference dependence theory suggests that customers' decision-making behavior is influenced by the difference between actual value and reference value (reference point) [8, 49]. The reference point can be divided into status quo reference and non-status quo reference [43]. The former refers to individuals using the current situation as a reference point, while the latter refers to situations where a non-objective status is used as a reference. Some scholars prefer to use minimum values, maximum values, or ideal points as reference points [9, 27]. Shi [30] proposed a cross-efficiency evaluation method based on prospect theory and using maximum profit as a reference point to analyze performance evaluation in organizational management. Liu and Popkowski Leszczyc [21] studied the reference price effect of historical price lists (provided in auction descriptions) on final prices. In queueing systems, Yang *et al.* [42] analyzed the optimal decisions of service providers based on price and waiting time reference points. Additionally, Terzi *et al.* [38] argue that the use of reference points exhibits clear consumer heterogeneity, meaning that different types of customers select different reference points.

Currently, research on reference theory is relatively well-developed, but there is limited literature that connects customer reference behavior with the observability of the service. Therefore, this study proposes that customer reference behavior can only exist when the EWS is observable. When the EWS is not observable, different types of customers cannot know each other's current status (status quo reference point), resulting in the disappearance of reference points. By combining customer reference behavior with the observability of the EWS, this study establishes the prerequisite conditions for the existence of reference behavior and enriches the application of reference theory in the operation of the EWS.

Pooling queues refer to organizing customers in waiting into a single queue served by a group of servers, while dedicated queues refer to each queue having its own dedicated server [7]. Most queueing research generally believes that operating with a pooled queue (*i.e.*, shared queue) rather than separate queues (*i.e.*, dedicated queues) is beneficial [6, 12, 32]. Benjaafar [6] identified the boundary conditions for the improvement of system efficiency with pooled queues. Gans *et al.* [12] confirmed that centralizing call centers in different regions can achieve higher operational efficiency. Through behavioral experiments, Rafaeli *et al.* [26] found that when queues are pooled, customers have stronger predictability and a higher level of fairness. This potential fairness difference is also one of the factors considered in this study.

However, pooling queues are not always advantageous [3, 51]. Some scholars have considered the adverse effects that merging queues may have due to changes in server-side factors. Rubinovitch [28] suggested that asymmetries in service rates may lead to inefficient sharing. Additionally, Gilbert and Weng [14] demonstrated that in competitive markets, pooling queues can lead to a decrease in capacity at equilibrium states. Song *et al.* [35] found that transitioning from a pooled queue system to a dedicated queue system in the emergency department can reduce patients' average waiting time and length of stay, further confirming the advantages of dedicated queues in certain scenarios. Furthermore, pooled queues may lead to a decrease in service quality. Wang *et al.* [41] analyzed the pooling effect in customer-intensive services and found that pooling agents tend to increase service rates to serve more customers, but this leads to a decrease in service quality. Sunar *et al.* [37] also demonstrated that even with the same servers and customers of the same type, pooled queues can result in worse performance than dedicated systems when delay-sensitive customers make joining decisions in an observable system.

Currently, these studies mainly consider the impact of service providers providing pooled queues or dedicated queues on system performance in different contexts. This study, based on the fairness of pooled queues considered by Rafaeli *et al.* [26] and using reference theory, analyzes customer reference behavior resulting from the adoption of dedicated queues, and examines the impact of pooled queues and dedicated queues on service providers' operations in EWS from a customer behavior perspective. This provides new explanations and support for whether service providers in EWS should adopt dedicated queues.

Additionally, our research is related to the literature on information sharing and delay announcements. Delay announcements are closely related to the experiential utility of customers in service systems and have received widespread attention from service providers and scholars in recent years [48, 50]. In practice, service providers often share delay information with customers in the form of delay announcements. In this regard, Bassamboo and Ibrahim [5] compared static and dynamic notifications to analyze when these announcements perform well.

Yuan *et al.* [45] considered the consumption behavior of customers under the influence of entertainment in the waiting area after service providers announced delay information and analyzed the competition and cooperation decisions among service providers accordingly. Furthermore, Singh *et al.* [31] also analyzed how such delay announcements affect the decisions and profits of service providers in a competitive environment where two service providers are vying for market share. Meanwhile, scholars such as Aksin *et al.* [2] and Jouini *et al.* [15] have empirically analyzed the impact of delay announcements on call center performance. In comparison to these studies, we distinguish delay announcements under different capacity usage modes, *i.e.*, when using pooled capacity, service providers announce uniform waiting times to different types of customers, whereas when using dedicated capacity, they announce different waiting times to different types of customers. By combining delay announcements with capacity usage modes, we have expanded upon this literature stream.

Built upon existing studies, our study advances the research on service operations management with the provision of extra value-added services in waiting areas. In particular, the contribution of our study can be summarized as follows: (1) Our study addresses the impact of the chargeable instead of free value-added services in waiting areas on service operational decisions such as pricing and capacity allocation. (2) We analyze the impact of the observability of the extra value-added services in waiting areas on operational decisions and system performance where the reference effect plays an important role in customers' value perception. (3) Considering capacity allocation in operating extra value-added services in waiting areas, we compare the advantages and disadvantages of pooled and dedicated capacities driven by the cost of capacity, where the results enrich our understanding of the joint impacts of customers' reference behavior, resource pooling, and delay information on service operations management. By filling the gap in the current literature, we aim to derive meaningful managerial insights and offer practical guidelines for service providers in operating extra value-added services in waiting areas.

3. Base model construction

3.1. Customer classification and utility

A service provider in the market offers base services to customers, and to alleviate customer's waiting anxiety, the service provider also offers Extra Waiting area value-added Services (EWS) options for customers to choose from. For example, airlines may offer VIP lounges for customers to purchase, and Chinese high-speed railway stations may provide business waiting rooms or VIP waiting rooms for customers to choose from. In the market, there are two types of customers: those who purchase EWS (E-type customers) and ordinary customers who do not purchase EWS. This also indicates that different customer types have corresponding consumption behaviors. This article assumes that customer categories are exogenous, depending on customer consumption habits and preferences, and in the long-term steady state, customer transfers can be considered non-existent [11]. The base arrival rates of E-type and O-type customers in the market are denoted as B_e and B_o, respectively. The delay sensitivity (unit time waiting cost) of both customer types is the same, and the unit time waiting cost for customers can be divided into opportunity cost and anxiety cost [16,46]. We represent the opportunity cost and anxiety cost per unit time for customers as t and h, respectively. EWS primarily reduces customers' perceived anxiety during waiting, and according to the study by Yuan *et al.* [45], when the level of EWS provided by the service provider is a, customers' anxiety sensitivity decreases to $he^{-\gamma a}$, where $\gamma > 0$. Table 1 shows the parameters and related definitions.

The service provider charges E-type and O-type customers p_e and p_o, respectively, and the customer's sensitivity to service prices is denoted as β. Therefore, when the customer's waiting time is w, the effective arrival rate for the two types of customers can be represented as follows:

$$\begin{cases} \lambda_e = B_e - tw - hwe^{-\gamma a} - \beta p_e, \\ \lambda_o = B_o - tw - hw - \beta p_o. \end{cases} \tag{1}$$

Given γ and a, the equation above indicates that the effective arrival rates of both E-type and O-type customers decrease linearly with increasing waiting time and price. This modeling assumption has been widely

Table 1. Parameters and definitions.

Parameters	Definitions
B_e or B_o	Base arrival rate for E-type or O-type customers
λ_e or λ_o	Effective arrival rate for E-type or O-type customers
ε_e or ε_o	Reference sensitivity for E-type or O-type customers
h	The cost of anxiety per unit time
t	Opportunity cost per unit of time
w	The waiting-time standard announced by the service provider
β	Customer sensitivity to price
γ	Customer sensitivity to EWS
p_e or p_o	The price charged to E-type customers or O-type customers
c	The unit cost required to maintain the service capacity μ
μ	Service capacity of service provider
η	The index of reference sensitivity differences among different customers
k	EWS prices
π	Service provider profit

applied in such studies [13,45]. It also demonstrates that the anxiety cost for E-type customers, after purchasing EWS, decreases to $he^{-\gamma a}$, while these customers need to pay a higher fee, p_e. Meanwhile, it is evident that a induces nonlinear effects on the anxiety cost for E-type customers. In the subsequent analysis, we will study the level of EWS and the extra prices that E-type customers need to pay.

3.2. Decision behavior of the service provider

When the service provider uses pooled capacity to serve two types of customers, the service provider can be considered as having only one server, and the service capacity μ follows a negative exponential distribution, which is in line with the classic assumption of the $M/M/1$ queuing model [13,45,47]. And we denoted c as the service cost to maintain unit capacity. In practice, the service provider is able to announce the waiting-time standard w [45], such as the boarding schedule at the airport or the waiting time display board at a restaurant. The average waiting time for the two types of customers in the pooled capacity service system is the same, which indicates that even though E-type customers purchase EWS, they do not reduce the actual waiting time, but only reduce the perceived cost of waiting. Just like customers paying for massage chair services provided by a cinema but cannot watch the movie in advance, they still need to wait with O-type customers. Therefore, according to the $M/M/1$ model, the service capacity μ can be represented as follows when w, λ_e, and λ_o are given:

$$\mu = \lambda_e + \lambda_o + \frac{1}{w}\cdot \tag{2}$$

According to the study conducted by Yuan *et al.* [45] and Mukhopadhyay *et al.* [24], it is known that the cost for the service provider to offer an EWS with level a is given by $C(a) = \frac{1}{2}\tau a^2$, where τ represents the cost coefficient. Therefore, the profit of the service provider can be represented as follows:

$$\pi = p_e\lambda_e + p_o\lambda_o - c\mu - C(a). \tag{3}$$

However, when the service provider uses dedicated capacity, it can be seen as having two independent service counters, providing services to E-type customers and O-type customers separately. Their service capacities are denoted as μ_e and μ_o, respectively. For customers of a specific type, they still receive service at a single service counter. The service provider specifies waiting-time standards for the two customer types as w_e and w_o. In other words, customers purchasing EWS may have the privilege of priority use of the service facilities. For example, Chinese high-speed train stations may offer priority boarding and ticket checking services to passengers who purchase the business waiting room, and most VIP airport lounges provide passengers with exclusive channels for ticket checking [10]. Therefore, the service capacities μ_e and μ_o can be expressed as follows:

$$\begin{cases} \mu_e = \lambda_e + \frac{1}{w_e}, \\ \mu_o = \lambda_o + \frac{1}{w_o}. \end{cases} \tag{4}$$

Therefore, the service provider profit function at this time can be expressed as follows:

$$\pi = p_e\lambda_e + p_o\lambda_o - c\left(\mu_e + \mu_o\right) - C(a). \tag{5}$$

The decision sequence for the service provider is as follows: first, determine the optimal level of EWS, denoted as a, and then, based on this service level a, formulate the optimal prices, p_e and p_o. We use a reverse solving approach for the optimization.

Yuan *et al.* [45] were the first to analyze the impact of waiting area entertainment provided by service providers on customers, exploring how waiting area entertainment interacts with pricing and capacity decisions. By examining scenarios of monopoly, competition, and coopetition, Yuan *et al.* [45] demonstrated the benefits of coopetition in service operations and provided managerial insights for operational execution and strategic interactions among firms under coopetition. Compared to Yuan *et al.* [45], our study has three main contributions:

(1) We categorized customers and studied the optimal operational decisions for service providers when offering differentiated services to different customers;
(2) We addressed the issue of whether service providers should offer observable EWS in practical operations;
(3) We investigated whether service providers should use pooled or dedicated capacity when offering EWS to different customers and provided theoretical support for this capacity mode selection.

4. Optimal decision with pooled capacity

When service providers use pooled capacity for service, the average waiting time for two types of customers is the same. For example, airlines announce boarding times and high-speed train companies publish departure schedules, and the waiting time for both types of customers is the same. In addition, service providers offer the option to purchase EWS and can set EWS as observable or unobservable. For example, when service providers use transparent tempered glass to separate the waiting areas for E-type customers and O-type customers, EWS is observable. Conversely, when they use non-transparent walls to separate these two waiting areas, EWS is unobservable. We first analyze the optimal price and EWS level in the unobservable scenario.

4.1. Optimal decisions in unobservable scenario

When EWS is unobservable, the two types of customers cannot perceive the current status of each other in the waiting area, so their effective demand functions are as shown in equation (1). Meanwhile, the service provider uses pooled capacity to serve customers, so the service capacity of the service provider is as shown in equation (2). By substituting equations (1) and (2) into the profit function equation (3), we can obtain the following profit function:

$$\pi = \left(p_e - c\right)\left(B_e - tw - hwe^{-\gamma a} - \beta p_e\right) + \left(p_o - c\right)\left(B_o - tw - hw - \beta p_o\right) - \frac{c}{w} - \frac{1}{2}\tau a^2. \tag{6}$$

Using the inverse solution method to solve the above formula, the following propositions can be obtained:

Proposition 1. *When $B_e \geq 2hw + \beta c$, there exists an optimal EWS intensity a_u^* that maximizes the service provider's profit. This a_u^* satisfies the following equation:*

$$\left(B_e - tw - hwe^{-\gamma a_u^*} - c\right)hw\gamma - 2\beta\tau a e^{\gamma a_u^*} = 0. \tag{7}$$

The pricing for E-type customers is $p_{u,e}^* = \frac{B_e - tw - hwe^{-\gamma a_u^*} + \beta c}{2\beta}$, *while the pricing for O-type customers is* $p_{u,o}^* = \frac{B_o - tw - hw + \beta c}{2\beta}$. *In this case, the optimal profit can be calculated as follows:*

$$\pi_u^* = \left(p_{u,e}^* - c\right)\left(B_e - tw - hwe^{-\gamma a_u^*} - \beta p_{u,e}^*\right) + \left(p_{u,o}^* - c\right)\left(B_o - tw - hw - \beta p_{u,o}^*\right) - \frac{c}{w} - \frac{1}{2}\tau a_u^{*2}. \quad (8)$$

The above proposition first restricts the base arrival rate B_e for E-type customers. Since EWS targets E-type customers, it is necessary to limit the lower bound of the base arrival rate for E-type customers. $B_e < 2hw + \beta c$ means that there are fewer E-type customers in the market, and offering EWS does not generate profits. This provides theoretical support from the perspective of base demand for whether service providers should offer EWS. We can draw the following conclusions based on the propositions above.

Proposition 1 also indicates that the pricing for E-type customers, $p_{u,e}^*$, increases with the increase of a_u^*, indicating that E-type customers need to spend more money to experience better EWS. However, the pricing for O-type customers, $p_{u,o}^*$, is not affected by a_u^*. Furthermore, both $p_{u,e}^*$ and $p_{u,o}^*$ increase with the increase of the unit service cost c, indicating that the service provider transfers the service cost c to both types of customers, resulting in a decrease in customer utility. The magnitude of the impact generated by EWS not only depends on a_u^*, but also on the sensitivity γ of E-type customers to a_u^*. Therefore, the following conclusions can be drawn.

Corollary 1. *As* γ *changes, we have:*

- *The EWS level* a_u^* *satisfies* $\lim_{\gamma\to 0} \frac{\partial a_u^*}{\partial \gamma} > 0$ *and* $\lim_{\gamma\to +\infty} \frac{\partial a_u^*}{\partial \gamma} < 0$, *indicating that* a_u^* *exhibits an increasing and then decreasing trend.*
- *The pricing for E-type customers satisfies* $\frac{\partial p_{u,e}^*}{\partial \gamma} > 0$, *while the pricing for O-type customers satisfies* $\frac{\partial p_{u,o}^*}{\partial \gamma} = 0$.
- *The profit of the service provider* π_u^* *satisfies* $\frac{\partial \pi_u^*}{\partial \gamma} > 0$.

The above corollary indicates that when customers have a low sensitivity γ towards EWS, service providers should offer higher quality EWS to attract customers. However, when γ is at a higher level, even a lower EWS level a_u^* can increase customer utility. Limited by the cost of EWS, $C(a) = \frac{1}{2}\tau a^2$, service providers should reduce a_u^* to minimize costs. Increasing γ will prompt service providers to increase the pricing for E-type customers, $p_{u,e}^*$. However, since EWS does not affect the utility of O-type customers, changes in γ do not affect $p_{u,o}^*$. The increase in $p_{u,e}^*$ further drives an increase in service provider profits, which is meaningful for the service provider. Therefore, many service providers advertise the EWS they offer to enhance the sensitivity γ of E-type customers, as seen in the promotion of VIP lounges by Airlines [1].

Based on the parameter settings of Yuan *et al.* [45] and Zhan *et al.* [47], the specific parameter settings are as follows: $B_e = 10$, $B_o = 6$, $h = 4$, $\beta = 1$, $r = 2$, $c = 1$, $\tau = 10$, $t = 3$, $\varepsilon_o = 2$, and $\varepsilon_e = 2$, which will be used consistently in the following text. Figure 1 illustrates the variation of the EWS level a_u^* with γ.

Figure 1 shows that as the customer's sensitivity γ to the EWS level increases, there is a turning point in the variation of the EWS level a_u^*. Furthermore, as the service provider increases the announced waiting-time standard w, not only does the service provider need to provide a higher level of a_u^* to compensate for the customer utility loss caused by the increase in w, but the turning point of a_u^* occurs earlier. This is because a larger a_u^* leads to higher service costs, and the service provider must reduce a_u^* as soon as possible to minimize cost investment.

In addition, based on Proposition 1, the separate price for the EWS, denoted as k_u^*, can be represented as follows:

$$k_u^* = p_{u,e}^* - p_{u,o}^* = \frac{B_e - B_o + hw\left(1 - e^{-\gamma a_u^*}\right)}{2\beta}. \quad (9)$$

In many cases, service providers sell the EWS separately. For example, some airlines sell tickets and VIP lounges separately. Therefore, having a separate price for the EWS benefits the service provider in improving operational flexibility[1].

[1]Other relevant conclusions regarding opportunity costs, anxiety costs, and wait-time standard in the waiting process can be found in the Appendixes A.1 and A.2.

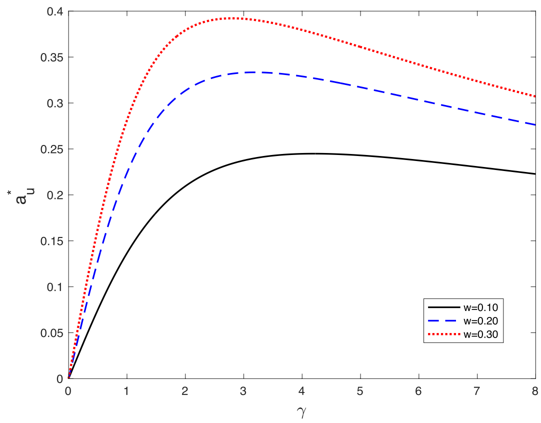

FIGURE 1. The variation of the EWS level a_u^* with γ.

4.2. Optimal decisions in observable situation

In the previous section, we analyzed the optimal decision-making and profit of service providers when providing unobservable EWS. However, in practice, some service providers make the EWS observable by using transparent tempered glass to separate two waiting areas, allowing customers of different types to see each other's waiting areas and perceive the EWS in different waiting areas. In this case, O-type customers perceive that the anxiety cost for E-type customers after enjoying the EWS becomes $he^{-\gamma a}w$. Therefore, with the anxiety cost of E-type customers as a reference point, O-type customers perceive the difference in anxiety cost between themselves and E-type customers as $hw\left(1-e^{-\gamma a}\right)$. According to the research by Liu *et al.* [22], O-type customers will have a perception of unfairness, and the decrease in demand caused by this perception of unfairness is $\varepsilon_o hw\left(1-e^{-\gamma a}\right)$, where ε_o is the reference sensitivity (loss sensitivity) of O-type customers. On the other hand, E-type customers will also compare their own anxiety cost with the anxiety cost of O-type customers, resulting in a sense of satisfaction. The increase in demand caused by this perception is $\varepsilon_e hw\left(1-e^{-\gamma a}\right)$, where ε_e is the reference sensitivity (gain sensitivity) of E-type customers. ε_e and ε_o indicate that E-type customers and O-type customers have different reference sensitivities, reflecting differences in customer reference behavior and aiding in our analysis of how customer reference behavior influences service providers' decisions and profits. Therefore, under observable EWS, the effective arrival rates of E-type customers and O-type customers can be rewritten as follows:

$$\begin{cases} \lambda_e = B_e - tw - hwe^{-\gamma a} - \beta p_e + \varepsilon_e hw\left(1-e^{-\gamma a}\right), \\ \lambda_o = B_o - tw - hw - \beta p_o - \varepsilon_o hw\left(1-e^{-\gamma a}\right). \end{cases} \tag{10}$$

And because $\mu = \lambda_e + \lambda_o + \frac{1}{w}$, equation (3) can be rewritten as follows:

$$\begin{aligned} \pi = &\left(p_e - c\right)\left(B_e - tw - hw_e e^{-\gamma a} - \beta p_e + \varepsilon_e hw\left(1-e^{-\gamma a}\right)\right) \\ &+ \left(p_o - c\right)\left(B_o - tw - hw_o - \beta p_o - \varepsilon_o h\left(1-e^{-\gamma a}\right)\right) - \frac{c}{w} - \frac{1}{2}\tau a^2. \end{aligned} \tag{11}$$

By optimizing the above profit function, the following proposition can be obtained.

Proposition 2. *When the base arrival rates of different types of customers satisfy* $\frac{(B_e - tw - 2hw - \varepsilon_e hw - \beta c)}{(B_o - tw - hw + \varepsilon_o hw - \beta c)} > \frac{\varepsilon_o}{1+\varepsilon_e}$*, there must exist an optimal EWS level* a_o^* *that satisfies*

$$\begin{aligned} &\left(B_e - tw - hwe^{-\gamma a_o^*} + \varepsilon_e hw\left(1-e^{-\gamma a_o^*}\right) - \beta c\right)\left(1+\varepsilon_e\right)hw\gamma \\ &- \left(B_o - tw - hw - \varepsilon_o hw\left(1-e^{-\gamma a_o^*}\right) - \beta c\right)\varepsilon_o hw\gamma - 2\beta\tau a_o^* e^{\delta a_o^*} = 0. \end{aligned} \tag{12}$$

Then, the optimal pricing for E-type customers and O-type customers can be calculated as follows:

$$p_{o,e}^{*} = \frac{B_e - tw - hwe^{-\gamma a_o^*} + \varepsilon_e hw\left(1 - e^{-\gamma a_o^*}\right) + \beta c}{2\beta},$$
$$p_{o,o}^{*} = \frac{B_o - tw - hw - \varepsilon_o hw\left(1 - e^{-\gamma a_o^*}\right) + \beta c}{2\beta}. \tag{13}$$

The optimal profit for the service provider can be calculated as follows:

$$\pi_o^* = \left(p_{o,e}^* - c\right)\left(B_e - tw - hwe^{-\gamma a_o^*} - \beta p_{o,e}^* + \varepsilon_e hw\left(1 - e^{-\gamma a_o^*}\right)\right)$$
$$+ \left(p_{o,o}^* - c\right)\left(B_o - tw - hw - \beta p_{o,o}^* - \varepsilon_o hw\left(1 - e^{-\gamma a_o^*}\right)\right) - \frac{c}{w} - \frac{1}{2}\tau a_o^{*2}. \tag{14}$$

Proposition 2 states that when ε_o and ε_e remain constant, the base arrival rates B_e of E-type customers and B_o of O-type customers need to satisfy $\frac{(B_e - tw - 2hw - \varepsilon_e hw - \beta c)}{(B_o - tw - hw + \varepsilon_o hw - \beta c)} > \frac{\varepsilon_o}{1+\varepsilon_e}$ for the service provider to obtain optimal profit by providing EWS. This indicates that the proportion of heterogeneous customer's base demand in the market determines whether the service provider should offer EWS. However, as ε_o increases, *i.e.*, when the loss sensitivity of O-type customers becomes stronger, it requires a higher B_e. Conversely, as ε_e increases, *i.e.*, when the gain sensitivity of E-type customers becomes stronger, it requires a higher B_o. This suggests that an increase in the reference sensitivity of one type of customer will increase the demand for the base arrival rate of the other type of customer.

Based on the above proposition, we can obtain the EWS price k_o^* as follows:

$$k_o^* = p_{o,e}^* - p_{o,o}^* = \frac{B_e - B_o + hw\left(1 - e^{-\gamma a_o^*}\right) + (\varepsilon_e + \varepsilon_o)\, hw\left(1 - e^{-\gamma a_o^*}\right)}{2\beta}. \tag{15}$$

Formula (15) illustrates that under observable EWS, the price k_o^* will be jointly influenced by ε_e and ε_o. Additionally, in contrast to Proposition 1, in this scenario, the O-type price $p_{o,o}^*$ will be affected by the EWS level. So, how will the reference behavior of customers of different types affect the service provider's decisions?

Corollary 2. (i) *As the sensitivity* ε_e *of E-type customers changes, it can be observed that* $\frac{\partial a_o^*}{\partial \varepsilon_e} > 0$, $\frac{\partial p_{o,e}^*}{\partial \varepsilon_e} > 0$, $\frac{\partial p_{o,o}^*}{\partial \varepsilon_e} < 0$, *and* $\frac{\partial \pi_o^*}{\partial \varepsilon_e} > 0$. (ii) *Conversely, as the loss aversion* ε_o *of O-type customers changes, it can be deduced that* $\frac{\partial a_o^*}{\partial \varepsilon_o} < 0$, $\frac{\partial p_{o,e}^*}{\partial \varepsilon_o} < 0$, $\frac{\partial \pi_o^*}{\partial \varepsilon_o} < 0$, *and* $\lim_{\varepsilon_o \to 0} \frac{\partial p_{o,o}^*}{\partial \varepsilon_o} < 0$, $\lim_{\varepsilon_o \to +\infty} \frac{\partial p_{o,o}^*}{\partial \varepsilon_o} > 0$. (iii) *Furthermore, it can also be concluded that* $\frac{\partial k_o^*}{\partial \varepsilon_e} > 0$ *and* $\frac{\partial k_o^*}{\partial \varepsilon_o} < 0$.

The above corollary indicates that as ε_e increases, service providers will increase the level of EWS and simultaneously increase the price for E-type customers while decreasing the price for O-type customers, which is beneficial for improving service providers' profits. This means that when E-type customers can obtain higher utility through reference behavior, service providers need to increase a_o^* to attract such customers and transfer service costs by charging higher $p_{o,e}^*$. As the increase in a_o^* leads to a stronger perception of unfairness for O-type customers in reference behavior, service providers must lower $p_{o,o}^*$ to attract such customers. In summary, the reference behavior of E-type customers is beneficial for improving service providers' profits. On the other hand, the impact of increasing ε_0 for O-type customers is opposite.

Furthermore, through the variation of the EWS price k_o^*, it can be observed that as ε_e increases, service providers will increase k_o^* to compensate for the cost increase caused by the rise in a_o^*. However, an increase in ε_0 will lead to a decrease in a_o^*, so service providers need to lower the EWS price k_o^*. Based on the above conclusions, further analysis can be conducted on the impact of ε_e, ε_o, and γ on the EWS level provided by service providers and the prices. In this case, specific parameter settings are $B_e = 10$, $B_o = 6$, $h = 4$, $\beta = 1$, $r = 2$, $c = 1$, $\tau = 10$, $t = 3$, $w = 0.10$, and $\varepsilon_o = 2$ or $\varepsilon_e = 2$. And the variation of EWS level and service

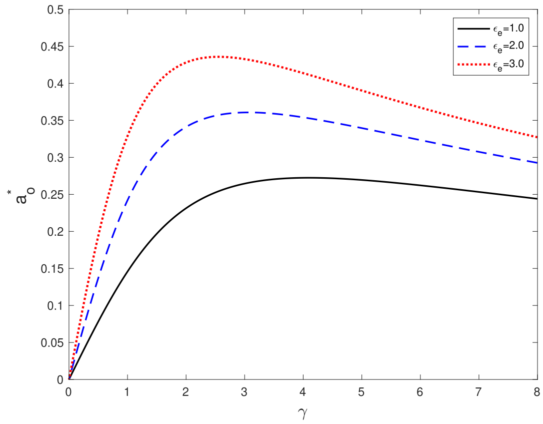

FIGURE 2. The impact of γ on a_o^* under different levels of ε_e.

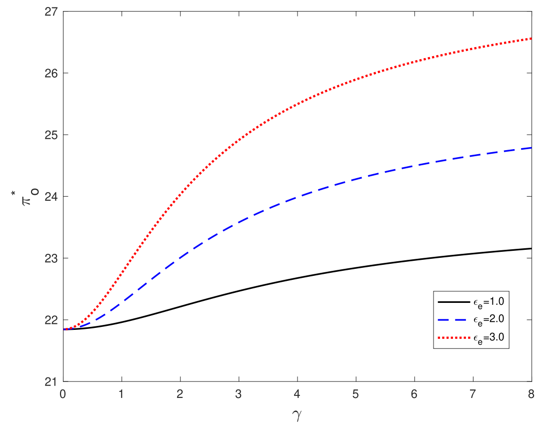

FIGURE 3. The impact of γ on π_o^* under different levels of ε_e.

providers' profits with respect to γ under different levels of heterogeneity in customer reference sensitivity is shown in Figures 2–5.

From Figures 2–5, it can be observed that as the sensitivity γ of E-type customers changes, the optimal EWS level a_o^* shows a trend of initially increasing and then decreasing, while the optimal service providers' profits π_o^* show an increasing trend. Additionally, as shown in Figure 2, it can be noticed that higher ε_e leads to a more significant change in a_o^* and a more pronounced increase in profits. This indicates that an increase in ε_e promotes the positive impact of γ on EWS level and profits. However, Figures 4–5 reveal that an increase in the reference sensitivity ε_o of O-type customers suppresses the positive impact of γ. Comparing Figures 2–5, it can be inferred that under the condition of the existence of the optimal solution, *i.e.*, when the base arrival rate of E-type customers is relatively high, the magnitude of change in the EWS level a_o^* and profits π_o^* with respect to a unit change in ε_e is greater than that with respect to a unit change in ε_o. In other words, a_o^* and π_o^* are more sensitive to changes in ε_e. Therefore, for service providers, the positive effects of increasing ε_e outweigh the effects of reducing ε_o. This implies that service providers should pay more attention to the reference behavior of E-type customers, which is beneficial for increasing profits.

Based on the above analysis, it can be concluded that the gain sensitivity of E-type customers and the loss sensitivity of O-type customers have opposite effects on service providers' profits. Therefore, we can propose a

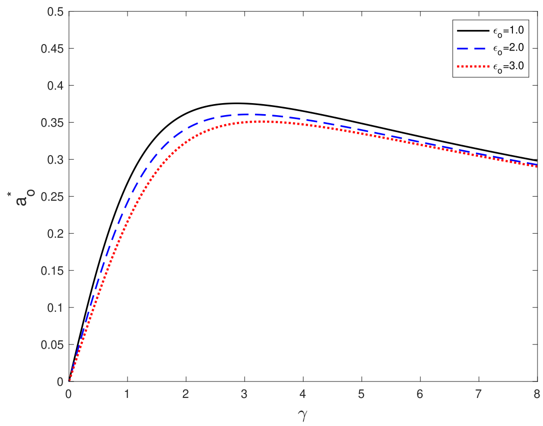

FIGURE 4. The impact of γ on a_o^* under different levels of ε_o.

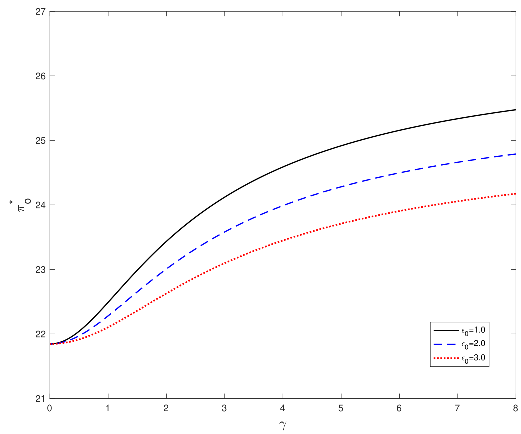

FIGURE 5. The impact of γ on π_o^* under different levels of ε_o.

special EWS operation mode, *i.e.*, by setting one-way observability to retain the acquisition sensitivity of E-type customers and eliminate the loss sensitivity of O-type customers. This EWS operation mode can be achieved by setting a one-way glass wall. According to the Appendix A.2, it can be concluded that this special operational mode can achieve higher profits by eliminating the loss sensitivity of O-type customers.

However, we must acknowledge the unethical nature of this operation mode, and one-way glass is often used in special occasions, with very limited application in public settings and a potential for non-compliance issues [34]. Therefore, we focus more on analyzing the decisions and profits under bidirectional observability and unobservability to determine the circumstances under which observable EWS should be implemented.

4.3. Observable *vs.* unobservable

In the observable case, customers of different types have reference effects on the service provider's decisions and profits based on their perceptions derived from the reference point. Therefore, the service provider can effectively enhance profits by suppressing negative effects and reinforcing positive effects. We introduce a new variable to aid in the subsequent analysis. Let η represent the degree of difference between the reference sensitivity ε_e of E-type customers and the reference sensitivity ε_o of O-type customers, and it is referred to as the Sensitivity

Differential Index. η can be expressed as follows:

$$\eta = \frac{\varepsilon_e - \varepsilon_o}{\varepsilon_o}. \tag{16}$$

η varies based on the characteristics of different industries. When O-type customers prioritize fairness more, for example in certain public services, $\eta < 0$. However, in industries that are primarily focused on customer experience and have a stronger social interaction effect, such as customized services, $\eta > 0$, meaning that E-type customers compare themselves to O-type customers to achieve a higher level of satisfaction. Therefore, η can be used to characterize the behavioral differences of heterogeneous customers regarding the EWS in various industries. To ensure that ε_o and ε_e are both zero only when EWS is unobservable, the equation can be rewritten as follows:

$$\varepsilon_e = (1+\eta)\varepsilon_o. \tag{17}$$

In the equation (17), $\eta > -1$. By using this equation, we can replace the customer reference sensitivity mentioned in Proposition 2. By comparing Propositions 1 and 2, there are two indifference points, η_a^P and η_π^P, respectively, which make the EWS level or profits equal in both cases. The indifference point for EWS level, η_a^P, satisfies the following equation:

$$\begin{gathered} hw\gamma\left[(1+\eta)\left(B_e - tw - hwe^{-\gamma a} + 2(1+\eta)\varepsilon_o hw\left(1-e^{-\gamma a}\right) - \beta c\right) + (1+\eta)hw\gamma\left(1-e^{-\gamma a}\right)\right] \\ -hw\gamma\left(B_o - tw - hw - 2\varepsilon_o hw\left(1-e^{-\gamma a}\right) - \beta c\right) = 0. \end{gathered} \tag{18}$$

And the indifference point for profits, η_π^P, satisfies the following equation:

$$\eta_\pi^P = \frac{\left(p_{o,o}^* - c\right)}{\left(p_{o,e}^* - c\right)} - 1. \tag{19}$$

And both indifference points satisfy $\eta_\pi^P > \eta_a^P > -1$. Then, we can get the following conclusion:

Corollary 3. *As η varies, there are several scenarios between the optimal decisions and profits in the observable and unobservable cases:*

- *When $-1 < \eta \leq \eta_a^P$, it is guaranteed that $a_o^* \leq a_u^*$, $p_{o,o}^* < p_{u,o}^*$ and $\pi_o^* < \pi_u^*$. Pricing for E-type customers and EWS pricing is uncertain.*
- *When $\eta_a^P < \eta \leq \eta_\pi^P$, it is guaranteed that $a_o^* > a_u^*$, $p_{o,e}^* > p_{u,e}^*$, $k_o^* > k_u^*$ and $\pi_o^* \leq \pi_u^*$. Pricing for O-type customers is uncertain.*
- *When $\eta > \eta_\pi^P$, it is guaranteed that $a_o^* > a_u^*$, $p_{o,e}^* > p_{u,e}^*$, $k_o^* > k_u^*$ and $\pi_o^* > \pi_u^*$. Pricing for O-type customers is uncertain.*

The above corollary provides choices between bidirectional observable EWS and unobservable EWS. While there is still some price uncertainty in each scenario, the level of EWS and the magnitude of profits are crucial for the service provider. The level of EWS determines the cost investment, and the relative size of profits directly affects the service provider's choices. When $-1 < \eta \leq \eta_a^P$, with a lower sensitivity differential index η, meaning ϵ_e is relatively low, the service provider should offer a lower EWS level. Meanwhile, the profits are lower than in the unobservable case. Therefore, the service provider should opt to forgo offering observable EWS.

When $\eta_a^P < \eta \leq \eta_\pi^P$, with a moderate η, even though the service provider can offer better EWS, set higher prices, they cannot achieve higher profits. Thus, the service provider should continue to offer unobservable EWS.

However, when η is relatively high, *i.e.*, $\eta > \eta_\pi^P$, not only can the service provider offer better EWS and set higher prices, but they can also earn higher profits. This is because the marginal revenue brought by E-type customers through reference behavior is higher than that of O-type customers. Therefore, the service provider can offer better EWS to attract E-type customers, increase prices to transfer service costs to customers, and thereby gain high profits. In this case, the optimal strategy for the service provider is to forgo unobservable EWS and choose observable EWS to strengthen customer reference behavior.

5. Optimal decision with dedicated capacity

In the previous sections, we analyzed the optimal decisions and profits for service providers when using pooled capacity for service. Using pooled capacity implies that service providers announce the same waiting-time standards w for both types of customers. However, in practice, some service providers may announce different waiting-time standards for different types of customers and utilize dedicated capacity to serve different types of customers. The variation in waiting-time standards will further impact the role of EWS and customer reference behavior in the observable case, which is the subject of our upcoming research. We will start by analyzing the unobservable case once again.

5.1. Optimal decisions in unobservable situation

When a service provider uses dedicated capacity for service, the waiting-time standards published by the service provider may vary due to differences in the service capacities for each type of customer. Let's assume that the waiting-time standards for E-type and O-type customers published by the service provider are denoted as w_e and w_o, respectively. If $w_e < w \leq w_o$, it means that E-type customers who have purchased EWS can receive faster service. First, it should be noted that in this unobservable scenario, the two types of customers can be considered as independent groups and cannot perceive each other's waiting times or their respective waiting areas. In this situation, the effective arrival rates for different types of customers are as follows:

$$\begin{cases} \lambda_e = B_e - tw_e - hw_e e^{-\gamma a} - \beta p_e, \\ \lambda_o = B_o - tw_o - hw_o - \beta p_o. \end{cases} \tag{20}$$

Based on the arrival rates and waiting-time standards for different types of customers, the service provider's capacity to serve different customers can be expressed as follows: $\mu_e = \lambda_e + \frac{1}{w_e}$, $\mu_o = \lambda_o + \frac{1}{w_o}\cdot$ Substituting equations (20) into (3) yields the following profit function:

$$\pi_u^D = (p_e - c)\left(B_e - tw_e - hw_e e^{-\gamma a} - \beta p_e\right) - \frac{c}{w_e} + (p_o - c)\left(B_o - tw_o - hw_o - \beta p_o\right) - \frac{c}{w_o} - \frac{1}{2}\tau a^2. \tag{21}$$

Through optimization, the following propositions can be obtained:

Proposition 3. *When $B_e - tw_e - 2hw_e - \beta c \geq 0$, there definitely exists an optimal EWS level a_u^D that maximizes service profits and satisfies*

$$\left(B_e - tw_e - hw_e e^{-\gamma a_u^D} - \beta c\right) hw_e \gamma - 2\beta\tau a e^{\gamma a_u^D} = 0. \tag{22}$$

Then, the optimal price for E-type customers is $p_{u,e}^D = \frac{B_e - tw_e - hw_e e^{-\gamma a_u^D} + \beta c}{2\beta}$, the optimal price for O-type customers is $p_{u,o}^D = \frac{B_o - tw_o - hw_o + \beta c}{2\beta}$, and the EWS price is $k_u^D = \frac{B_e - B_o + t(w - w_e) + h\left(w - w_e e^{-\gamma a}\right)}{2\beta}\cdot$ The service provider's optimal profit can be calculated as follows:

$$\begin{aligned} \pi_u^D = {} & \left(p_{u,e}^D - c\right)\left(B_e - tw_e - hw_e e^{-\gamma a_u^D} - \beta p_{u,e}^D\right) - \frac{c}{w_e} \\ & + \left(p_{u,o}^D - c\right)\left(B_o - tw_o - hw_o - \beta p_{u,o}^D\right) - \frac{c}{w_o} - \frac{1}{2}\tau (a_u^D)^2. \end{aligned} \tag{23}$$

Proposition 3 provides the requirement for the base arrival rate B_e for E-type customers when using dedicated capacity. Compared to Proposition 1, because $w_e < w$, indicating shorter waiting-time standards for E-type customers published by the service provider, the lower limit requirement for B_e will be lower. This means that a lower B_e is needed to encourage the service provider to offer higher-quality EWS.

5.2. Optimal decisions in observable situation

When EWS is observable, and the service provider uses dedicated capacity, O-type customers not only see that E-type customers are enjoying better EWS but also notice that the service provider is providing faster waiting-time standards for E-type customers. As a result, different types of customers exhibit reference behavior towards both anxiety costs and opportunity costs during the waiting process. Their effective arrival rates can be expressed as follows:

$$\begin{cases} \lambda_e = B_e - tw_e - hw_e e^{-\gamma a} + \varepsilon_e \left(t\left(w - w_e\right) + h\left(w - w_e e^{-\gamma a}\right)\right) - \beta p_e, \\ \lambda_o = B_o - tw - hw - \varepsilon_o \left(t\left(w - w_e\right) + h\left(w - w_e e^{-\gamma a}\right)\right) - \beta p_o. \end{cases} \tag{24}$$

Since $t\left(w - w_e\right)$ and $h\left(w - w_e e^{-\gamma a}\right)$ represent the differences in opportunity costs and anxiety costs for the two types of customers, the term $t\left(w - w_e\right) + h\left(w - w_e e^{-\gamma a}\right)$ in the equation indicates that customers not only compare anxiety costs during the waiting process but also compare opportunity costs. Moreover, because $\mu_e = \lambda_e + \frac{1}{w_e}$ and $\mu_o = \lambda_o + \frac{1}{w}$, the service provider's profit function is $\pi = \left(p_e - c\right)\lambda_e + \left(p_o - c\right)\lambda_o - \frac{c}{w_e} - \frac{c}{w} - \frac{1}{2}\tau a^2$. Solving this profit function through optimization leads to the following proposition.

Proposition 4. *When the base arrival rates of two customer types satisfy*

$$\frac{B_e - tw_e - 2\left(1 + \varepsilon_e\right)hw_e + \varepsilon_e\left(t\left(w - w_e\right) + hw\right) - \beta c - \left(1 + \varepsilon_e\right)tw_e}{B_o - tw - hw + 2\varepsilon_o hw_e - \varepsilon_o\left[t\left(w - w_e\right) + hw\right] - \beta c + \varepsilon_o tw_e} \geq \frac{\varepsilon_o}{1 + \varepsilon_e},$$

there exists an optimal EWS level a_o^D such that the service provider's profit is optimized, and a_o^D satisfies the following equation:

$$\begin{aligned} &\left(B_e - tw_e - hw_e e^{-\gamma a_o^D} + \varepsilon_e\left(t\left(w - w_e\right) + h\left(w - w_e e^{-\gamma a_o^D}\right)\right) - \beta c\right)\left(1 + \varepsilon_e\right)hw_e\gamma \\ &\quad - \left(B_o - tw - hw - \varepsilon_o\left(t\left(w - w_e\right) + h\left(w - w_e e^{-\gamma a_o^D}\right)\right) + \beta c\right)\varepsilon_o hw_e\gamma - 2\beta\tau a_o^D e^{\gamma a_o^D} = 0. \end{aligned} \tag{25}$$

Therefore, the charges for E-type customers $p_{o,e}^D$, the charges for O-type customers $p_{o,o}^D$, the price of EWS k_o^D, and the service provider's profits π_o^D can be expressed as follows:

$$\begin{aligned} p_{o,e}^D &= \frac{B_e - tw_e - hw_e e^{-\gamma a_o^D} + \varepsilon_e\left(t\left(w - w_e\right) + h\left(w - w_e e^{-\gamma a_o^D}\right)\right) + \beta c}{2\beta}, \\ p_{o,o}^D &= \frac{B_o - tw - hw - \varepsilon_o\left(t\left(w - w_e\right) + h\left(w - w_e e^{-\gamma a_o^D}\right)\right) + \beta c}{2\beta}, \\ k_o^D &= \frac{B_e - B_o + \left(1 + \varepsilon_e + \varepsilon_o\right)\left(t\left(w - w_e\right) + h\left(w - w_e e^{-\gamma a_o^D}\right)\right)}{2\beta}, \\ \pi_o^D &= \left(p_{o,e}^D - c\right)\lambda_e + \left(p_{o,o}^D - c\right)\lambda_o - \frac{c}{w_e} - \frac{c}{w} - \frac{1}{2}\tau\left(a_o^D\right)^2. \end{aligned} \tag{26}$$

Similar to the prerequisites for other decisions, Proposition 4 first provides the conditions for the base arrival rates of different customer types under observable EWS. It is evident that as the service provider shortens the waiting-time standard w_e for E-type customers, the requirement for the base arrival rate B_e of E-type customers in the market will decrease. This means that, compared to the service provider using pooled capacity, in this scenario, the service provider is more inclined to provide EWS.

In pooled capacity, there is no difference in waiting-time standards between different customer types, so there is no perception of differences in opportunity costs during the waiting process. However, in dedicated capacity,

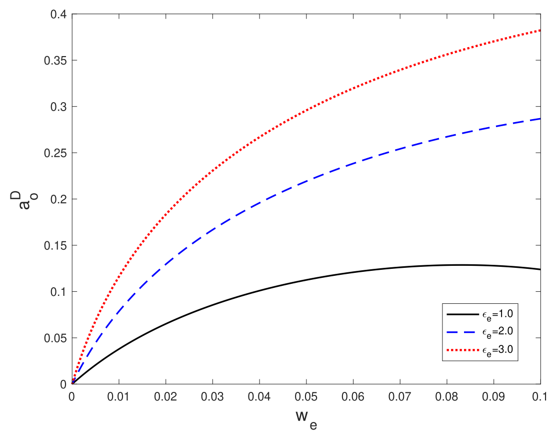

FIGURE 6. The impact of w_e on a_o^D under different levels of ε_e.

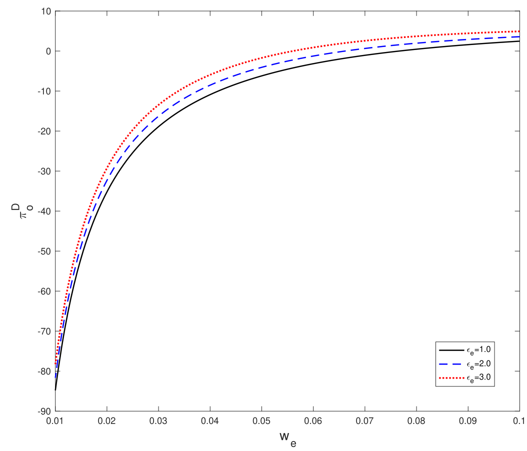

FIGURE 7. The impact of w_e on π_o^D under different levels of ε_e.

different customer types will exhibit different behaviors due to differences in waiting-time standards. The impact of the waiting-time standard can be found in Corollary A.4 of Appendix A.2.

Figures 6–9 illustrate the variations in EWS levels and profit as w_e changes with variations in ε_e and ε_o.

Figures 6 and 8 illustrate that a_o^D follows an opposite trend with w_e as ε_e and ε_o vary. With increasing ε_o, the service provider's a_o^D provided using dedicated capacity should increase. When ε_e is relatively small, a_o^D doesn't always increase with the increase of w_e but exhibits a trend of first increasing and then decreasing. This is because the increase in w_e implies a decrease in θ, which confirms Corollary A.4. On the other hand, a_o^D should decrease as ε_o increases, and when ε_o is relatively large, a_o^D shows a trend of first increasing and then decreasing with the increase in w_e.

Figures 7 and 9 depict the variations in profit π_o^D, illustrating that profit increases with the service provider's announced waiting-time standard w_e. However, at lower values of w_e, it is possible for the service provider to incur negative profit. This is because a smaller w_e implies that the service provider needs to invest more in service costs. Despite the simultaneous increase in customer demand, the revenue increment from the increased demand cannot offset the losses caused by the increased service costs. Furthermore, the opposing effects of ε_e and ε_o are also reflected in the service provider's profit. Additionally, similar to Figures 2–5, in conditions where

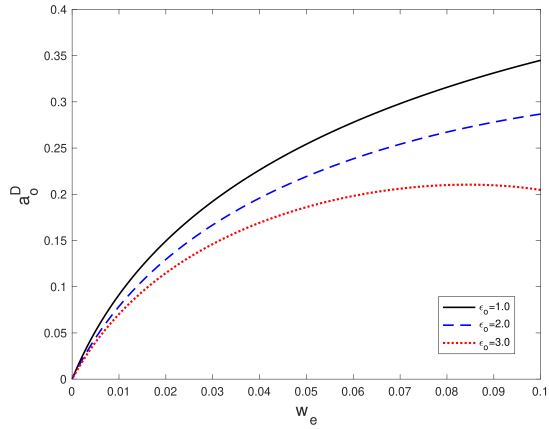

FIGURE 8. The impact of w_e on a_o^D under different levels of ε_o.

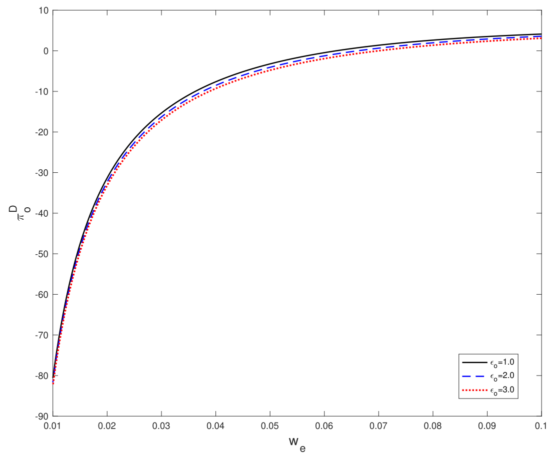

FIGURE 9. The impact of w_e on π_o^D under different levels of ε_o.

optimal solutions exist, especially when the base arrival rate of E-type customers is relatively high, the EWS level a_o^D and profit π_o^D exhibit more significant variations with unit changes in ε_e than with unit changes in ε_o. Based on this finding, service providers should place greater emphasis on enhancing the reference sensitivity of E-type customers. For example, offering services and products with a more conspicuous nature, such as trending foods and popular photo spots, can be advantageous for service providers in achieving higher profits by providing observable EWS.

5.3. Observable *vs.* unobservable

Comparing the optimal decisions and profits in the unobservable and observable cases under dedicated capacity, there exist two indifference points η_a^D and η_π^D, which respectively make the EWS level and profit equal in both scenarios. And both indifference points satisfy $\eta_\pi^D > \eta_a^D > -1$.

Corollary 4. *As η varies, the following scenarios exist between the optimal decisions and profits in the observable and unobservable cases under dedicated capacity:*

– *When* $-1 < \eta \leq \eta_a^D$*, it is guaranteed that* $a_o^D \leq a_u^D$*,* $p_{o,o}^D < p_{u,o}^D$ *and* $\pi_o^D < \pi_u^D$*. Pricing for E-type customers and EWS pricing is uncertain.*
– *When* $\eta_a^D < \eta \leq \eta_\pi^D$*, it is guaranteed that* $a_o^D > a_u^D$*,* $p_{o,e}^D > p_{u,e}^D$*,* $k_o^D > k_u^D$ *and* $\pi_o^D \leq \pi_u^D$*. Pricing for O-type customers is uncertain.*
– *When* $\eta > \eta_\pi^D$*, it is guaranteed that* $a_o^D > a_u^D$*,* $p_{o,e}^D > p_{u,e}^D$*,* $k_o^D > k_u^D$ *and* $\pi_o^D > \pi_u^D$*. Pricing for O-type customers is uncertain.*

The above corollary is similar to Corollary 3, as both illustrate the impact of the heterogeneity in sensitivity levels between E-type and O-type customers on the observability of EWS. Combining Corollaries 3 and 4, it is evident that in both dedicated capacity and pooled capacity settings, when the sensitivity difference index is relatively low, *i.e.*, ε_e is relatively low, providing unobservable EWS can result in higher profits for the service provider. Additionally, in this scenario, E-type customers can experience a higher level of EWS.

However, when the sensitivity difference index is moderate, *i.e.*, $\eta_a^D < \eta \leq \eta_\pi^D$, providing unobservable EWS can still yield profits higher than providing observable EWS, but the EWS level experienced by E-type customers is relatively lower. As ε_e gradually increases, when the sensitivity difference index between customers is relatively high, offering observable EWS not only results in higher profits for the service provider but also enhances the level of EWS. Therefore, when ε_e is at a high level, the optimal strategy for the service provider is to provide observable EWS to reinforce customer behavioral responses, thus achieving higher profits.

6. Pooled capacity *vs.* dedicated capacity

6.1. Comparison in unobservable scenarios

In Sections 4 and 5, we respectively determine the optimal decisions, and profits for service providers under pooled capacity and dedicated capacity. The question of whether service providers offering EWS should adopt pooled capacity or dedicated capacity is a crucial one, and one that we will focus on next. First, we turn our attention to the comparison in unobservable scenarios. According to Propositions 1 and 3, the level of EWS in unobservable scenarios satisfies the following equation:

$$\left(B_e - tw - hwe^{-\gamma a_u^*} - c\right) hw\gamma - 2\beta\tau a e^{\gamma a_u^*} = 0,$$
$$\left(B_e - tw_e - hw_e e^{-\gamma a_u^D} - \beta c\right) hw_e\gamma - 2\beta\tau a e^{\gamma a_u^D} = 0.$$

According to Propositions 1 and 3, the profits under pooled capacity and dedicated capacity are respectively:

$$\pi_u^* = \left(p_{u,e}^* - c\right)\left(B_e - tw - hwe^{-\gamma a_u^*} - \beta p_{u,e}^*\right) + \left(p_{u,o}^* - c\right)\left(B_o - tw - hw - \beta p_{u,o}^*\right) - \frac{c}{w} - \frac{1}{2}\tau\left(a_u^*\right)^2,$$
$$\pi_u^D = \left(p_{u,e}^D - c\right)\left(B_e - tw_e - hw_e e^{-\gamma a_u^D} - \beta p_{u,e}^D\right) - \frac{c}{w_e} + \left(p_{u,o}^D - c\right)\left(B_o - tw_o - hw_o - \beta p_{u,o}^D\right) - \frac{c}{w_o} - \frac{1}{2}\tau(a_u^D)^2.$$

By comparing the two formulas above, we get the following conclusions.

Corollary 5. *The level of EWS under dedicated capacity is lower than the level of EWS under pooled capacity, i.e.,* $a_u^D < a_u^*$*.*

Corollary 6. *There exists* $\bar{c}$ *such that* $(p_{u,e}^D - \bar{c})(t + he^{-\gamma a_u^D}) = \frac{\bar{c}}{w_e^2}$*. When* $c \leq \bar{c}$*, we have* $\pi_u^D \geq \pi_u^*$*. However, when* $c > \bar{c}$*, we have* $\pi_u^D < \pi_u^*$*.*

Corollaries 5 and 6 suggest that when the service provider uses dedicated capacities to offer shorter waiting times for customers who have purchased unobservable EWS, the level of EWS tends to decrease. This is because the actual waiting time for E-type customers is shortened compared to the case of pooled capacity. The service provider doesn't need to provide a better EWS to bring higher utility to customers. However, when the service provider offers observable EWS, the EWS level in dedicated capacity will be lower than in pooled capacity under the same conditions only when the base arrival rate of E-type customers is high. This is because having more E-type customers means the service provider doesn't have to provide a higher EWS level and can reduce costs by lowering the EWS level.

6.2. Comparison in observable scenarios

Next, we compare the EWS levels and the profits of service providers under different capacity usage modes in the observable EWS scenario. According to Propositions 2 and 4, the EWS levels under pooled capacity and dedicated capacity are respectively:

$$\begin{aligned}&\left(B_e - tw - hwe^{-\gamma a_o^*} + \varepsilon_e hw\left(1 - e^{-\gamma a_o^*}\right) - \beta c\right)(1+\varepsilon_e)\,hw\gamma \\ &\quad - \left(B_o - tw - hw - \varepsilon_o hw\left(1 - e^{-\gamma a_o^*}\right) - \beta c\right)\varepsilon_o hw\gamma - 2\beta\tau a_o^* e^{\delta a_o^*} = 0,\\ &\left(B_e - tw_e - hw_e e^{-\gamma a_o^D} + \varepsilon_e\left(t(w-w_e) + h\left(w - w_e e^{-\gamma a_o^D}\right)\right) - \beta c\right)(1+\varepsilon_e)\,hw_e\gamma \\ &\quad - \left(B_o - tw - hw - \varepsilon_o\left(t(w-w_e) + h\left(w - w_e e^{-\gamma a_o^D}\right)\right) + \beta c\right)\varepsilon_o hw_e\gamma - 2\beta\tau a_o^D e^{\gamma a_o^D} = 0.\end{aligned}$$

According to Propositions 2 and 4, the profits under pooled capacity and dedicated capacity are respectively:

$$\begin{aligned}\pi_o^* &= \left(p_{o,e}^* - c\right)\lambda_e + \left(p_{o,o}^* - c\right)\lambda_o - \frac{c}{w} - \frac{1}{2}\tau(a_o^*)^2,\\ \pi_o^D &= \left(p_{o,e}^D - c\right)\lambda_e + \left(p_{o,o}^D - c\right)\lambda_o - \frac{c}{w_e} - \frac{c}{w} - \frac{1}{2}\tau(a_o^D)^2.\end{aligned}$$

From the comparison of the above formulas, we can draw the following Corollaries.

Corollary 7. *When EWS is observable and $\bar{H}_e > \widetilde{H}_e \geq \frac{\varepsilon_o}{1+\varepsilon_e}$, we find $a_o^D < a_o^*$. Conversely, we can get $a_o^D > a_o^*$. $\bar{H}_e$ and $\widetilde{H}_e$ are illustrated in Corollary A.4.*

Corollary 8. *When $c \leq \tilde{c}$, we have $\frac{\partial \pi_o^D}{\partial w_e} \geq 0$, which leads to $\pi_o^D \geq \pi_o^*$. However, when $c > \tilde{c}$, we have $\frac{\partial \pi_o^D}{\partial w_e} < 0$, resulting in $\pi_o^D < \pi_o^*$. Here, $\tilde{c}$ satisfies $\left(p_{o,e}^D - \tilde{c}\right)\left[t + he^{-\gamma a} + \varepsilon_e t + \varepsilon_e he^{-\gamma a}\right] = \left(p_{o,o}^D - \tilde{c}\right)\left[\varepsilon_o t + \varepsilon_o he^{-\gamma a}\right] + \frac{\tilde{c}}{(w-\theta)^2}$.*

Combining Corollaries 5–8, we can conclude that whether dedicated capacity can yield higher profits than pooled capacity depends on the relative value of the unit service cost c, regardless of whether EWS is observable. When c is relatively low, the service provider can achieve higher profits using dedicated capacity because the lower c results in lower additional costs associated with the use of a dedicated service channel. In this case, the cost reduction from the decrease in EWS level compensate for the service cost. Conversely, when c is relatively high, using dedicated capacity may lead to a rapid increase in service costs. In such cases, using pooled capacity becomes the optimal strategy for the service provider. This also implies that the cost advantage is a key factor in using dedicated capacity. Therefore, service providers must evaluate the current service costs when considering whether to adopt dedicated capacity.

7. Conclusion

7.1. Findings

We have constructed a mathematical model using queueing theory to analyze the decision-making related to EWS operations when using pooled capacity and dedicated capacity. We have also examined the impact of EWS observability on decisions change. Based on the theoretical model analysis and numerical simulation, the following conclusions have been obtained:

- The optimal level of EWS shows a trend of first increasing and then decreasing as customer sensitivity to it rises. When customer sensitivity to EWS is low, providers should offer a higher level of EWS to attract customers. Conversely, when sensitivity is high, providers should reduce the EWS level to minimize costs. This implies that linearly increasing the EWS level may not be beneficial for profit growth.
- The heterogeneity of customer reference sensitivity will affect the service provider's choice of EWS observability. When the reference sensitivity of E-type customers is higher than that of O-type customers, meaning a higher level of heterogeneity in customer reference sensitivity, observable EWS generates more significant value through customer reference behavior. Therefore, the service provider should provide observable EWS to enhance the reference behavior of heterogeneous customers and increase profits. Conversely, when the heterogeneity in reference sensitivity between E-type customers and O-type customers is lower, the service provider should provide unobservable EWS.
- Under unobservability, the EWS level under dedicated capacity is always lower than that under pooled capacity. However, once EWS is observable, the EWS level under dedicated capacity is lower than that under pooled capacity only when the base arrival rate of E-type customers is high. When the service provider uses dedicated capacity to provide unobservable EWS for different types of customers, the actual waiting time for E-type customers is shorter compared to using pooled capacity. Therefore, the service provider does not need to provide better EWS to bring more utility to customers. This also illustrates that when the service provider offers unobservable EWS, pooled capacity provides better waiting area services for customers.
- The key factor in determining whether the service provider uses dedicated capacity to serve different customers is the service cost. When the service cost is relatively low, the service provider can earn higher profits by using dedicated capacity because the reduced EWS level and increased prices result in higher unit revenue. Once the service cost is relatively high, then using pooled capacity is a better strategy, providing theoretical support for whether the service provider adopts dedicated capacity.

7.2. Management inspiration

We conducted an in-depth study on the observability of EWS operations and the using mode of capacity. From this research, we have gained the following key management insights.

First, when considering whether to provide observable EWS, it's important to take into account the heterogeneity of customer reference sensitivity levels. Service providers can conduct market research to understand the reference sensitivity of different types of customers. When the heterogeneity in reference sensitivity between E-type and O-type customers is relatively high, service providers can enhance the perceived value for E-type customers by offering observable EWS and increasing prices. This is beneficial for increasing profits. However, when the heterogeneity in reference sensitivity is low, and O-type customers have a stronger loss aversion, providing observable EWS will lead to a rapid decrease in demand among O-type customers, which is unfavorable for profit growth. In such cases, service providers should offer unobservable EWS to mitigate customer reference behavior.

Second, whether to use pooled capacity to serve different types of customers needs to consider the impact of service costs. When service costs are low, the cost advantage of using dedicated capacity outweighs the pooling effect under pooled capacity. Therefore, service providers need to use dedicated capacity to serve different types of customers and provide shorter waiting-time standards to E-type customers. However, once service costs are higher, using pooled capacity becomes the superior strategy.

For example, in the hotel industry, hotel operators can provide better value-added services in the waiting area for VIP customers who are waiting to check in, such as a pleasant waiting environment and complimentary beverages, to enhance their utility. Moreover, when service costs are relatively low, hotel staff should provide dedicated service to these VIP customers, further enhancing their value. Furthermore, when the gain sense from enjoying EWS by VIP customers significantly outweighs the loss sense by ordinary customers, hospitality practitioners should provide observable waiting area services. These insights have practical implications for operations in industries such as airlines, airport operations, and railway stations.

7.3. Deficiencies and future prospects

In our research, we analyzed the impact of EWS observability on customer consumption based on reference effects. We assumed that customer types are not interchangeable, but in the long-term operation, customer consumption behaviors may change, leading to a transformation in the market's customer structure. This is a question worth exploring in the future. Additionally, we believe that service capacity can be adjusted based on waiting-time standards, such as through temporary staffing. However, in practice, once service capacity becomes limited, it's necessary to analyze the allocation of capacity across different channels. This can assist service providers in addressing operational decisions when faced with limited service capacity and is an area of focus for future research.

Appendix A. Some auxiliary conclusions

A.1. The impact of opportunity cost and anxiety cost

EWS mainly affects the anxiety of customers during the waiting process, but cannot reduce the opportunity cost of waiting. With changes in t and h, we can draw the following conclusions.

Corollary A.1. (i) *As the customer's unit opportunity cost t changes, we can have* $\frac{\partial a_u^*}{\partial t} < 0$, $\frac{\partial p_{u,e}^*}{\partial t} < 0$, $\frac{\partial p_{u,o}^*}{\partial t} < 0$, *and* $\frac{\partial \pi_u^*}{\partial t} < 0$.

(ii) *However, as the customer's unit anxiety cost h changes, we can have* $\frac{\partial a_u^*}{\partial h} < 0$, $\frac{\partial p_{u,o}^*}{\partial h} < 0$, $\frac{\partial \pi_u^*}{\partial h} < 0$, *and* $\lim_{h\to 0} \frac{\partial p_{u,e}^*}{\partial h} < 0$, $\lim_{h\to +\infty} \frac{\partial p_{u,e}^*}{\partial h} > 0$.

Proof. Given $F = (B_e - tw - hwe^{-\gamma a_u^*} - \beta c)hw\gamma - 2\beta\tau a e^{\gamma a_u^*}$, the derivative of F with respect to t is $\frac{\partial F}{\partial t} = -hw^2\gamma < 0$. Therefore, $\frac{\partial a_u^*}{\partial t} = -\frac{\partial F/\partial t}{\partial F/\partial a} < 0$. Similarly, $\frac{\partial p_{u,e}^*}{\partial t} = \frac{-w+hw\gamma e^{-\gamma a_u^*}(\partial a_u^*/\partial t)}{2\beta} < 0$ and $\frac{\partial p_{u,o}^*}{\partial t} = \frac{-t}{2\beta} < 0$. Based on the envelope theorem, $\frac{\partial \pi_u^*}{\partial t} = -w(p_{u,e}^* - c) - w(p_{u,o}^* - c) < 0$.

The derivative of F with respect to h is $\frac{\partial F}{\partial h} = (B_e - tw - 2hwe^{-\gamma a_u^*} - \beta c)w\gamma > 0$. Therefore, $\frac{\partial a_u^*}{\partial h} = -\frac{\partial F/\partial h}{\partial F/\partial a} > 0$.

For $p_{u,e}^* = \frac{B_e - tw - hwe^{-\gamma a_u^*} + \beta c}{2\beta}$, the derivative with respect to h is $\frac{\partial p_{u,e}^*}{\partial h} = \frac{-we^{-\gamma a_u^*} + h\gamma we^{-\gamma a_u^*}(\partial a_u^*/\partial h)}{2\beta} = \frac{we^{-\gamma a_u^*}(h\gamma(\partial a_u^*/\partial h)-1)}{2\beta}$. Therefore, $\lim_{h\to 0}(\frac{\partial p_{u,e}^*}{\partial h}) < 0$ and $\lim_{h\to +\infty}(\frac{\partial p_{u,e}^*}{\partial h}) > 0$.

For $p_{u,o}^* = \frac{B_o - tw - hw + \beta c}{2\beta}$, the derivative with respect to h is $\frac{\partial p_{u,o}^*}{\partial h} = \frac{-w}{2\beta} < 0$. Based on the envelope theorem, the derivative of the profit function π_u^* with respect to h is $\frac{\partial \pi_u^*}{\partial h} = -(p_{u,e}^* - c)we^{-\gamma a_u^*} - (p_{u,o}^* - c)w < 0$. □

It is unexpected that an increase in unit opportunity cost t leads to a decrease in a_u^*. It is generally believed that when the customer's waiting cost increases, the service provider should provide a better EWS to correspondingly reduce the customer's waiting cost. However, our research indicates that opportunity cost and anxiety cost have opposite effects on the EWS level. This is because the EWS cannot reduce the increase in opportunity cost, so the service provider can only ensure customer utility by lowering prices. The decrease in price requires the service provider to lower the EWS level to save on cost expenditure, thus ensuring optimal profit. However, when the anxiety cost h increases, the service provider can reduce customer anxiety by increasing a_u^*, while $p_{u,e}^*$ shows a decreasing-then-increasing trend. This indicates that when h is relatively low, the service provider can further reduce $p_{u,e}^*$ to attract E-type customers. However, once h is relatively high, the service provider must increase $p_{u,e}^*$ to compensate for the cost increase caused by the increase in a_u^*. Furthermore, further reduction in $p_{u,o}^*$ will result in a decrease in the service provider's optimal profit. Overall, changes in both customer opportunity cost and anxiety cost lead to a decrease in O-type prices and service provider profits, reflecting the adverse impact of increased customer costs on profits. The variation of these two costs has opposite effects on the level of EWS.

A.2. The impact of waiting time in all cases

A.2.1. The unobservable scenario with pooled capacity

Corollary A.2. *As the waiting-time standard w announced by the service provider increases, it can be shown that $\frac{\partial a_u^*}{\partial w} > 0$, $\frac{\partial k_u^*}{\partial w} > 0$, and $\frac{\partial \pi_u^*}{\partial w} < 0$.*

Corollary A.2 indicates that when the service provider announces a longer waiting-time standard w, the service provider should provide a higher quality EWS and compensate for the service cost of the EWS by setting a higher price k. However, a higher k does not effectively increase the service provider's profit. This is because the decrease in profit is mainly influenced by the decrease in the effective arrival rate of two types of customers. When w is longer, both types of customers realize that they will have to wait longer in the system. Although the service provider will provide a better EWS to alleviate the waiting anxiety of E-type customers, the inevitable increase in waiting time directly leads to a rise in costs. Moreover, for O-type customers, the increase in w directly increases the waiting cost, further reducing customer demand and lowering profits. This provides a management insight that in the absence of observable EWS, the way for service providers to increase profits is to announce shorter waiting-time standards as much as possible, while simultaneously reducing the optimal level and price of EWS.

With the development of internet technology, modern service providers are able to inform customers of wait times through electronic display boards, mobile terminals, and other means [45, 50]. Therefore, based on the managerial insights mentioned above, in some high-end service industries such as private medical clinics, upscale restaurants, and luxury retail stores, customer experience is often emphasized, including minimizing customer wait times as much as possible. Consequently, these service providers may be willing to invest more resources to provide more efficient and expedited services, thereby increasing customer demand and generating more revenue for themselves.

A.2.2. The unobservable scenario with dedicated capacity

In the case of using dedicated capacity with unobservable EWS, we can derive the following conclusions about the changes in waiting-time standards:

Corollary A.3. (i) *As the waiting-time standard w_e decreases, the EWS level a_u^D will decrease. However, as w_o increases, a_u^D remains fixed.*

(ii) *When $(p_{u,e}^D - c)(t + he^{-\gamma a_u^D}) < \frac{c}{w_e^2}$, we have $\frac{\partial \pi_u^D}{\partial w_e} > 0$; and when $(p_{u,e}^D - c)(t + he^{-\gamma a_u^D}) > \frac{c}{w_e^2}$, then we have $\frac{\partial \pi_u^D}{\partial w_e} < 0$. Similarly, when $(p_{u,o}^D - c)(t + h) < \frac{c}{w_o^2}$, we have $\frac{\partial \pi_u^D}{\partial w_o} > 0$; and when $(p_{u,o}^D - c)(t + h) > \frac{c}{w_o^2}$, then we have $\frac{\partial \pi_u^D}{\partial w_o} < 0$.*

The above corollary indicates that under dedicated capacity, when the service provider discloses a lower waiting-time standard w_e to the E-type customers, the service provider should offer a lower level of EWS, but w_o does not affect the formulation of EWS. As for the service provider's profit, when both w_o and w_e are relatively small, the profit π_u^D increases with the increase of w_o and w_e. However, w_o and w_e cannot be infinitely small, otherwise it will lead to negative profit for the service provider. Conversely, the profit π_u^D decreases with the increase of w_o and w_e. This also indicates that larger profit and smaller w_o and w_e are not conducive to profit increase. Because smaller waiting-time standards require the service provider to incur higher service costs, while larger waiting-time standards will lead to a decrease in customer demand. For the purpose of simplifying subsequent analysis without loss of generality, we assume that the waiting-time standard disclosed by the service provider to O-type customers is consistent with that in pooled capacity, *i.e.*, $w_o = w$. Therefore, we can obtain $w_e < w$, which will be used in the subsequent text.

A.2.3. The observable scenario with dedicated capacity

In the case of using dedicated capacity with observable EWS, for the sake of subsequent analysis, let's assume that the reduction in waiting-time standard w_e for E-type customers compared to w is θ, *i.e.*, $w - w_e = \theta$. We can then derive the following corollary.

Corollary A.4. *With the variation of θ, we can derive the following:*

– *If $\bar{H}_e > \widetilde{H}_e \geq \frac{\varepsilon_o}{1+\varepsilon_e}$, we have $\frac{\partial a_o^D}{\partial \theta} < 0$; Once $\bar{H}_e \geq \frac{\varepsilon_o}{1+\varepsilon_e} > \widetilde{H}_e$, we can have $\frac{\partial a_o^D}{\partial \theta} > 0$, where*

$$\bar{H}_e = \frac{B_e - tw_e - 2(1+\varepsilon_e)hw_e + \varepsilon_e[t(w-w_e)+hw] - \beta c}{B_o - tw - hw + 2\varepsilon_o hw_e - \varepsilon_o[t(w-w_e)+hw] - \beta c},$$
$$\widetilde{H}_e = \frac{B_e - tw_e - 2(1+\varepsilon_e)hw_e + \varepsilon_e[t(w-w_e)+hw] - \beta c - (1+\varepsilon_e)tw_e}{B_o - tw - hw + 2\varepsilon_o hw_e - \varepsilon_o[t(w-w_e)+hw] - \beta c + \varepsilon_o tw_e}.$$

– *There exists a $\tilde{c}$ satisfying*

$$\left(p_{o,e}^D - \tilde{c}\right)\left[t + he^{-\gamma a} + \varepsilon_e t + \varepsilon_e he^{-\gamma a}\right] = \left(p_{o,o}^D - \tilde{c}\right)\left[\varepsilon_o t + \varepsilon_o he^{-\gamma a}\right] + \frac{\tilde{c}}{(w-\theta)^2}.$$

When $c \leq \tilde{c}$, we have $\frac{\partial \pi_o^D}{\partial \theta} \geq 0$, while when $c > \tilde{c}$, there exists $\frac{\partial \pi_o^D}{\partial \theta} < 0$.

Corollary A.4 first indicates that when other conditions are fixed, if the base arrival rate B_e of E-type customers is high, then as θ increases, service providers offering observable EWS should gradually reduce the EWS level to decrease service cost expenses. This is because an increase in θ implies that E-type customers can derive higher perceived utility from the reduction in waiting time. Therefore, the service provider no longer needs to offer a higher EWS level to reduce the actual waiting costs for customers. However, when B_e is relatively low, even though an increase in θ can provide customers with increased utility, the service provider still needs to offer a higher level of EWS to attract more customers.

As for the service provides' profit, when θ gradually increases, the demand for E-type customers in the system increases due to shorter waiting times. Hence, only when the service cost c is relatively low can the increased revenue from the demand offset the service cost investment, leading to an increase in the service provider's profit. However, if the service cost c is high, the cost increase caused by the demand increment due to the increase in θ may harm the service provider's profit.

A.3. Single-sided observability

Proposition A.1. *When the base arrival rates of different types of customers satisfy $B_e - tw - 2hw - \varepsilon_e hw - \beta c \geq 0$, there must exist an optimal EWS level a_o^{1*} that satisfies*

$$\left(B_e - tw - hwe^{-\gamma a_o^{1*}} + \varepsilon_e hw\left(1 - e^{-\gamma a_o^{1*}}\right) - \beta c\right)(1+\varepsilon_e)hw\gamma - 2\beta\tau a_o^{1*} e^{\delta a_o^{1*}} = 0. \tag{A.1}$$

(The superscript "1" indicates the one-way observable scenario)

Corollary A.5. *Based on Proposition A.1, the optimal pricing for E-type customers and O-type customers are*

$$p_{o,e}^{1*} = \frac{B_e - tw - hwe^{-\gamma a_o^{1*}} + \varepsilon_e hw\left(1 - e^{-\gamma a_o^{1*}}\right) + \beta c}{2\beta},$$
$$p_{o,o}^{1*} = \frac{B_o - tw - hw + \beta c}{2\beta}.$$

The optimal profit for the service provider is

$$\begin{aligned}\pi_o^{1*} = &\left(p_{o,e}^{1*} - c\right)\left(B_e - tw - hwe^{-\gamma a_o^{1*}} - \beta p_{o,e}^{1*} + \varepsilon_e hw\left(1 - e^{-\gamma a_o^{1*}}\right)\right)\\ &+ \left(p_{o,o}^{1*} - c\right)\left(B_o - tw - hw - \beta p_{o,o}^{1*}\right) - \frac{c}{w} - \frac{1}{2}\tau a_o^{1*2}.\end{aligned} \tag{A.2}$$

Proof. When there is no loss aversion for O-type customers, the effective arrival rates for both customer types can be expressed as follows:

$$\begin{cases}\lambda_e = B_e - tw - hw_e e^{-\gamma a} - \beta p_e + \varepsilon_e hw(1-e^{-\gamma a}),\\ \lambda_o = B_o - tw - hw_o - \beta p_o.\end{cases}$$

Therefore, the service provider's profit function is given by

$$\pi_o^1 = (p_e - c)\left(B_e - tw - hw_e e^{-\gamma a} - \beta p_e + \varepsilon_e hw\left(1-e^{-\gamma a}\right)\right) + (p_o - c)(B_o - tw - hw_o - \beta p_o) - \frac{c}{w} - \frac{1}{2}\tau a^2.$$

Taking derivatives with respect to p_e and p_o, we obtain

$$\frac{\partial \pi_o^1}{\partial p_e} = B_e - tw - hw_e e^{-\gamma a} - 2\beta p_e + \varepsilon_e hw\left(1-e^{-\gamma a}\right) + \beta c,$$

and

$$\frac{\partial \pi_o^1}{\partial p_o} = B_o - tw - hw_o - 2\beta p_o.$$

Thus, the optimal prices are given by

$$p_{o,e}^{1*} = \frac{B_e - tw - hwe^{-\gamma a} + \varepsilon_e hw(1-e^{-\gamma a}) + \beta c}{2\beta},$$

and

$$p_{o,o}^{1*} = \frac{B_o - tw - hw + \beta c}{2\beta}.$$

Substituting these into the profit function yields

$$\pi_o^1 = \frac{\left(B_e - tw - hwe^{-\gamma a} + \varepsilon_e hw\left(1-e^{-\gamma a}\right) - \beta c\right)^2}{4\beta} + \frac{\left(B_o - tw - hw - \beta c\right)^2}{4\beta} - \frac{1}{2}\tau a^2,$$

and differentiating with respect to a gives

$$\frac{\partial \pi}{\partial a} = \frac{\left(B_e - tw - hwe^{-\gamma a} + \varepsilon_e hw\left(1-e^{-\gamma a}\right) - \beta c\right)(1+\varepsilon_e)hw\gamma e^{-\gamma a}}{2\beta} - \tau a,$$

and the second derivative is

$$\frac{\partial^2 \pi}{\partial a^2} = -\frac{\left(B_e - tw - 2(1+\varepsilon_e)hwe^{-\gamma a} + \varepsilon_e hw - \beta c\right)(1+\varepsilon_e)hw\gamma^2 e^{-\gamma a}}{2\beta} - \tau.$$

Therefore, when $B_e - tw - 2hw - \varepsilon_e hw - \beta c \geq 0$, there exists an optimal EWS intensity a_o^{1*} satisfying

$$\left(B_e - tw - hwe^{-\gamma a_o^{1*}} + \varepsilon_e hw\left(1-e^{-\gamma a_o^{1*}}\right) - \beta c\right)(1+\varepsilon_e)hw\gamma - 2\beta\tau a_o^{1*} e^{\delta a_o^{1*}} = 0.$$

The optimal prices and profit at this point are given by

$$p_{o,e}^{1*} = \frac{B_e - tw - hwe^{-\gamma a_o^{1*}} + \varepsilon_e hw(1-e^{-\gamma a_o^{1*}}) + \beta c}{2\beta},$$
$$p_{o,o}^{1*} = \frac{B_o - tw - hw + \beta c}{2\beta},$$

and

$$\begin{aligned}\pi_o^{1*} = &\left(p_{o,e}^{1*} - c\right)\left(B_e - tw - hwe^{-\gamma a_o^{1*}} - \beta p_{o,e}^{1*} + \varepsilon_e hw\left(1-e^{-\gamma a_o^{1*}}\right)\right)\\ &+ \left(p_{o,o}^{1*} - c\right)\left(B_o - tw - hw - \beta p_{o,o}^{1*}\right) - \frac{c}{w} - \frac{1}{2}\tau {a_o^{1*}}^2.\end{aligned} \tag{A.3}$$

□

Compared to Propositions 1 and A.1 suggests that if only E-type customers can observe the waiting area where O-type customers are located, then the minimum base arrival rate of E-type customers required by the service provider to provide EWS will be higher. In other words, only when there are more E-type customers will the service provider choose to provide EWS.

Corollary A.6. *When service providers use one-way observability, it is certain that* $a_o^{1*} > a_u^*$*,* $p_{o,e}^{1*} > p_{u,e}^*$*,* $p_{o,o}^{1*} = p_{u,o}^*$*,* $\pi_o^{*1} > \pi_u^*$*.*

Proof. In the case of one-way observability, the optimal EWS level a_o^{1*} satisfies

$$\left(B_e - tw - hwe^{-\gamma a_o^{1*}} + \varepsilon_e hw\left(1 - e^{-\gamma a_o^{1*}}\right) - \beta c\right)(1+\varepsilon_e)\, hw\gamma - 2\beta\tau a_o^{1*} e^{\delta a_o^{1*}} = 0.$$

Let $F_3 = (B_e - tw - hwe^{-\gamma a_o^{1*}} + \varepsilon_e hw(1 - e^{-\gamma a_o^{1*}}) - \beta c)(1+\varepsilon_e)hw\gamma - 2\beta\tau a_o^{1*} e^{\delta a_o^{1*}}$. The derivative of F_3 with respect to ε_e is given by

$$\frac{\partial F_3}{\partial \varepsilon_e} = \left(B_e - tw - hwe^{-\gamma a_o^{1*}} + (1+2\varepsilon_e)\, hw\left(1 - e^{-\gamma a_o^{1*}}\right) - \beta c\right) hw\gamma > 0.$$

Thus, $\frac{\partial a_o^{1*}}{\partial \varepsilon_e} > 0$, indicating that there exists a positive relationship between ε_e and a_o^{1*}. When $\varepsilon_e = 0$, $a_o^{1*} = a_u^*$, and therefore, $a_o^{1*} > a_u^*$. Given that $p_{(o,e)}^{1*} = \frac{B_e - tw - hwe^{-\gamma a_o^{1*}} + \varepsilon_e hw(1-e^{-\gamma a_o^{1*}}) + \beta c}{2\beta}$, the derivative of $p_{(o,e)}^{1*}$ with respect to ε_e is

$$\frac{\partial p_{o,e}^{1*}}{\partial \varepsilon_e} = \frac{\left(hw - hwe^{-\gamma a_o^{1*}} + (1+\varepsilon_e)\, hwe^{-\gamma a_o^{1*}}\right)\frac{\partial a_o^{1*}}{\partial \varepsilon_e}}{2\beta} > 0,$$

implying that $p_{o,e}^{1*} > p_{u,e}^*$. Simultaneously, $p_{o,o}^{1*} = p_{u,o}^*$. The profit function π_o^{1*} is given by

$$\begin{aligned}\pi_o^{1*} &= \left(p_{o,e}^{1*} - c\right)\left(B_e - tw - hw_e e^{-\gamma a_o^{1*}} - \beta p_{o,e}^{1*} + \varepsilon_e hw\left(1 - e^{-\gamma a_o^{1*}}\right)\right)\\ &\quad + \left(p_{o,o}^{1*} - c\right)\left(B_o - tw - hw_o - \beta p_{o,o}^{1*}\right) - \frac{c}{w} - \frac{1}{2}\tau a_o^{1*2}.\end{aligned}$$

Applying the envelope theorem to ε_e yields

$$\frac{\partial \pi_o^{1*}}{\partial \varepsilon_e} = \left(p_{o,e}^{1*} - c\right) hw\left(1 - e^{-\gamma a_o^{1*}}\right) > 0,$$

and when $\varepsilon_e = 0$, $\pi_o^{1*} = \pi_u^*$, indicating that $\pi_o^{1*} > \pi_u^*$. ☐

The above corollary confirms that in the case of one-way observability, the EWS level provided by service providers is higher than in the unobservable scenario, and the price for E-type customers is also higher. Since $p_{o,e}^{1*} > p_{u,e}^*$ and $p_{o,o}^{1*} = p_{u,o}^*$, it is evident that $k_o^{1*} > k_u^*$, indicating that the EWS price under one-way observability is higher than in the unobservable scenario. More importantly, π_o^{1*} is higher than π_u^*, which means that providing one-way observable EWS can yield higher profits, providing theoretical support for service providers to set observability for EWS. However, we must acknowledge the unethical nature of this operation mode, and one-way glass is often used in special occasions, with very limited application in public settings and a potential for non-compliance issues [34]. Therefore, we focus more on analyzing the decisions and profits under bidirectional observability and unobservability to determine the circumstances under which observable EWS should be implemented.

Appendix B. Proof of propositions and corollaries

B.1. Proof of Proposition 1

We know that the profit of the service provider is given by

$$\pi = (p_e - c)\left(B_e - tw - hwe^{-\gamma a} - \beta p_e\right) + (p_o - c)\left(B_o - tw - hw - \beta p_o\right) - \frac{c}{w} - \frac{1}{2}\tau a^2. \tag{B.1}$$

Taking the partial derivatives of π with respect to p_e and p_o, we obtain $\frac{\partial \pi}{\partial p_e} = (B_e - hwe^{-\gamma a} - 2\beta p_e + \beta c)$ and $\frac{\partial \pi}{\partial p_o} = B_o - hw - 2\beta p_o + \beta c$. Setting $\frac{\partial \pi}{\partial p_e} = 0$ and $\frac{\partial \pi}{\partial p_o} = 0$, we get $p_{u,e} = \frac{B_e - hwe^{-\gamma a} + \beta c}{2\beta}$ and $p_{u,o} = \frac{B_o - hw + \beta c}{2\beta}$. Substituting these into the profit function, we have $\pi = \frac{(B_e - hwe^{-\gamma a} - \beta c)^2}{4\beta} + \frac{(B_o - hw - \beta c)^2}{4\beta} - \frac{1}{2}\tau a^2$. Taking the first and second derivatives of π with respect to a, we obtain $\frac{\partial \pi}{\partial a} = \frac{(B_e - hwe^{-\gamma a} - \beta c)hw\gamma e^{-\gamma a}}{2\beta} - \tau a$ and $\frac{\partial^2 \pi}{\partial a^2} = \frac{-hw\gamma(B_e - 2hwe^{-\gamma a} - \beta c)\gamma e^{-\gamma a}}{2\beta} - \tau$. Therefore, when $B_e \geq 2hw + \beta c$, there must exist $\frac{\partial^2 \pi}{\partial a^2} < 0$, indicating the existence of an optimal EWS level a_u^* satisfying $(B_e - hwe^{-\gamma a_u^*} - c)hw\delta - 2\beta\tau a e^{\gamma a_u^*} = 0$, and optimizing the service provider's profit. The optimal prices under this condition are $p_{u,e}^* = \frac{B_e - tw - hwe^{-\gamma a_u^*} + \beta c}{2\beta}$ and $p_{u,o}^* = \frac{B_o - tw - hw + \beta c}{2\beta}$. Therefore, the profit is

$$\pi_u^* = \left(p_{u,e}^* - c\right)\left(B_e - tw - hwe^{-\gamma a_u^*} - \beta p_{u,e}^*\right) + \left(p_{u,o}^* - c\right)\left(B_o - tw - hw - \beta p_{u,o}^*\right) - \frac{1}{2}\tau a_u^*. \tag{B.2}$$

B.2. Proof of Corollary 1

Let $F(a) = (B_e - tw - hwe^{-\gamma a} - \beta c)hw\gamma - 2\beta\tau a e^{\gamma a}$. Taking the first and second derivatives of $F(a)$ with respect to a, we obtain $F_a = \frac{\partial F(a)}{\partial a} = (hw\gamma)^2 e^{-\gamma a} - 2\beta\tau e^{\gamma a} - 2\beta\tau a\gamma e^{\gamma a}$, $F_{aa} = -\gamma(hw\gamma)^2 e^{-\delta a} - 4\beta\tau\gamma e^{\gamma a} - 2\beta\tau a\gamma^2 e^{\gamma a} < 0$. $F_a(\tilde{a}) = 0$, meaning that the first derivative is equal to 0 at $a = \tilde{a}$, and since $F_{aa} < 0$, this implies that $F(\tilde{a})$ has a maximum value at $a = \tilde{a}$. As a increases, at $a = a_2^*$, $F = 0$, so $a_u^* > \tilde{a}$, and for F_a, it must be the case that $F_a(a_u^*) < 0$.

Taking the derivative of F with respect to γ yields the following function:

$$\frac{\partial F(a)}{\partial \gamma} = \left(B_e - tw - hwe^{-\gamma a} - \beta c\right)hw + (hw)^2 a\gamma e^{-\gamma a} - 2\beta\tau a^2 e^{\gamma a}. \tag{B.3}$$

For the above equation, we have $\lim_{\gamma\to 0} F(a) = (B_e - tw - hw - \beta c)hw\gamma - 2\beta\tau a e^{\gamma a}$, which gives $\lim_{\gamma\to 0}\frac{\partial F}{\partial \gamma} = (B_e - tw - hwe^{-\gamma a} - \beta c)hw - 2\beta\tau a^2$. Since $\lim_{\delta\to 0} F(a) = 0$ at $a = a_u^*$, it follows that $\lim_{\gamma\to 0}\frac{\partial F}{\partial \gamma} = (B_e - tw - hwe^{-\gamma a} - \beta c)hw > 0$. Additionally, we can obtain $\lim_{\gamma\to +\infty}\frac{\partial F(a)}{\partial \gamma} = (B_e - tw - hwe^{-\gamma a} - \beta c)hw - 2\beta\tau a^2 e^{\gamma a} < 0$, which leads to $\lim_{\gamma\to 0}\frac{\partial a_u^*}{\partial \gamma} = -\frac{\frac{\partial F}{\partial \gamma}}{\frac{\partial F}{\partial a_u^*}} > 0$ and $\lim_{\gamma\to +\infty}\frac{\partial a_u^*}{\partial \gamma} = -\frac{\frac{\partial F}{\partial \gamma}}{\frac{\partial F}{\partial a_u^*}} < 0$.

$p_{u,e}^* = \frac{B_e - tw - hwe^{-\gamma a_u^*} + \beta c}{2\beta}$, and taking the derivative with respect to γ , we can obtain the following formula:

$$\frac{\partial p_{u,e}^*}{\partial \gamma} = \frac{hwe^{-\gamma a_u^*}}{2\beta}\left(\frac{-a_u^* 2\beta\tau e^{\gamma a_u^*} - \gamma\left[\left(B_e - tw - hwe^{-\gamma a_u^*} - \beta c\right)hw\right]}{\frac{\partial F}{\partial a_u^*}}\right) > 0.$$

By the envelope theorem, the profit function π_u^* with respect to γ yields $\frac{\partial \pi_u^*}{\partial \gamma} = \left(p_{u,e}^* - c\right)a_u^* hwe^{-\gamma a_u^*} > 0$.

B.3. Proof of Corollary A.2

Given $F = (B_e - hwe^{-\gamma a} - \beta c)hw\gamma - 2\beta\tau a e^{\gamma a}$, the derivative of F with respect to w is $\frac{\partial F}{\partial w} = h\delta(B_e - 2hwe^{-\delta a_u^*} - \beta c) > 0$. Thus, we can obtain $\frac{\partial a_u^*}{\partial w} = -\frac{\partial F/\partial w}{\partial F/\partial a} > 0$.

Given the individual EWS prices $k_u^* = p_{u,e}^* - p_{u,o}^* = \frac{B_e - B_o + hw(1-e^{-\gamma a_u^*})}{2\beta}$, the derivative with respect to w is $\frac{\partial k}{\partial w} = \frac{h(1-e^{-\gamma a_u^*}) + hw\gamma e^{-\gamma a_u^*}(\partial a_u^*/\partial w)}{2\beta} > 0$. Additionally, based on the envelope theorem, $\frac{\partial \pi_u^*}{\partial w} = -(p_{u,e}^* - c)he^{-\gamma a_u^*} - h(p_{u,o}^* - c) < 0$.

B.4. Proof of Proposition 2

Given the profit function $\pi = (p_e - c)(B_e - tw - hw_e e^{-\gamma a} - \beta p_e + \varepsilon_e hw(1-e^{-\gamma a})) + (p_o - c)(B_o - tw - hw_o - \beta p_o - \varepsilon_o h(1-e^{-\gamma a})) - \frac{c}{w} - \frac{1}{2}\tau a^2$, the partial derivatives with respect to p_e and p_o are $\frac{\partial \pi}{\partial p_e} = B_e - tw - hw_e e^{-\gamma a} - 2\beta p_e + \varepsilon_e hw(1-e^{-\gamma a}) + \beta c$ and $\frac{\partial \pi}{\partial p_o} = B_o - tw - hw_o - 2\beta p_o - \varepsilon_o h(1-e^{-\gamma a}) + \beta c$. Setting $\frac{\partial \pi}{\partial p_e} = 0$ and $\frac{\partial \pi}{\partial p_o} = 0$, we obtain $p_{o,e} = \frac{B_e - tw - hwe^{-\gamma a} + \varepsilon_e hw(1-e^{-\gamma a}) + \beta c}{2\beta}$ and $p_{o,o} = \frac{B_o - tw - hw - \varepsilon_o hw(1-e^{-\gamma a}) + \beta c}{2\beta}$. Substituting these into the profit function, we get $\pi_o^* = \frac{(B_e - tw - hwe^{-\gamma a} + \varepsilon_e hw(1-e^{-\gamma a}) - \beta c)^2}{4\beta} + \frac{(B_o - tw - hw - \varepsilon_o hw(1-e^{-\gamma a}) - \beta c)^2}{4\beta} - \frac{1}{2}\tau a^2$. The first and second derivatives of π_o^* with respect to a are $\frac{\partial \pi_o^*}{\partial a} = \frac{(B_e - tw - hwe^{-\gamma a} + \varepsilon_e hw(1-e^{-\gamma a}) - \beta c)(1+\varepsilon_e)hw\gamma e^{-\gamma a}}{2\beta} - \frac{(B_o - tw - hw - \varepsilon_o hw(1-e^{-\gamma a}) - \beta c)\varepsilon_o hw\gamma e^{-\gamma a}}{2\beta} - \tau a$ and $\frac{\partial^2 \pi_o^*}{\partial a^2} = -\frac{(B_e - tw - 2(1+\varepsilon_e)hwe^{-\gamma a} + \varepsilon_e hw - \beta c)(1+\varepsilon_e)hw\gamma^2 e^{-\gamma a}}{2\beta} + \frac{(B_o - tw - hw - \varepsilon_o hw + 2\varepsilon_o hwe^{-\gamma a} - \beta c)\varepsilon_o hw\gamma^2 e^{-\gamma a}}{2\beta} - \tau$. Since $\frac{\partial^2 \pi_o^*}{\partial a^2} < -\frac{(B_e - tw - 2hw - \varepsilon_e hw - \beta c)(1+\varepsilon_e)}{2\beta} + \frac{(B_o - tw - hw + \varepsilon_o hw - \beta c)\varepsilon_o}{2\beta} - \tau$, it follows that when $\frac{(B_e - tw - 2hw - \varepsilon_e hw - \beta c)}{(B_o - tw - hw + \varepsilon_o hw - \beta c)} > \frac{\varepsilon_o}{1+\varepsilon_e}$, there exists $\partial^2 \pi_o^*/\partial a^2 < 0$, and there exists an optimal EWS level a_o^* such that the profit of the service provider is maximized, and a_o^* satisfies the following equation: $(B_e - tw - hwe^{-\gamma a_o^*} + \varepsilon_e hw(1-e^{-\gamma a_o^*}) - \beta c)(1+\varepsilon_e)hw\gamma - (B_o - tw - hw - \varepsilon_o hw(1-e^{-\gamma a_o^*}) - \beta c)\varepsilon_o hw\gamma - 2\beta\tau a_o^* e^{\delta a_o^*} = 0$. Therefore, the optimal prices and profit at this time are: $p_{o,e}^* = \frac{B_e - tw - hwe^{-\gamma a_o^*} + \varepsilon_e hw(1-e^{-\gamma a_o^*}) + \beta c}{2\beta}$, $p_{o,o}^* = \frac{B_o - tw - hw - \varepsilon_o hw(1-e^{-\gamma a_o^*}) + \beta c}{2\beta}$, and $\pi_o^* = (p_{o,e}^* - c)(B_e - tw - hw_e e^{-\gamma a_o^*} - \beta p_{o,e}^* + \varepsilon_e hw(1-e^{-\gamma a_o^*})) + (p_{o,o}^* - c)(B_o - tw - hw_o - \beta p_{o,o}^* - \varepsilon_o hw(1-e^{-\gamma a_o^*})) - \frac{1}{2}\tau a_o^2$.

B.5. Proof of Corollary 2

(i) Differentiating F with respect to ε_e yields $\frac{\partial F}{\partial \varepsilon_e} = hw\gamma(B_e - tw - 2(1+\varepsilon_e)hwe^{-\gamma a} + \varepsilon_e hw - c) > 0$. Therefore, it is certain that $\frac{\partial a_o^*}{\partial \varepsilon_e} > 0$. Knowing that $p_{o,e}^* = \frac{B_e - tw + \varepsilon_e hw - (1+\varepsilon_e)hwe^{-\gamma a_o^*} + \beta c}{2\beta}$, differentiating with respect to ε_e yields $\frac{\partial p_{o,e}^*}{\partial \varepsilon_e} = \frac{hw - hwe^{-\gamma a} + (1+\varepsilon_e)hwe^{-\gamma a}\frac{\partial a_o^*}{\partial \varepsilon_e}}{2\beta} > 0$; and knowing $p_{o,o}^* = \frac{B_o - tw - hw - \varepsilon_o hw(1-e^{-\gamma a_o^*}) + \beta c}{2\beta}$, differentiating with respect to ε_e yields $\frac{\partial p_{o,o}^*}{\partial \varepsilon_e} = (-\varepsilon_o hw\gamma e^{-\gamma a_o^*}\frac{\partial a_o^*}{\partial \varepsilon_e})/2\beta < 0$. Knowing that $k_o^* = p_{o,e}^* - p_{o,o}^* = \frac{B_e - B_o + hw(1-e^{-\gamma a_o^*}) + (\varepsilon_e + \varepsilon_o)hw(1-e^{-\gamma a_o^*})}{2\beta}$, differentiating with respect to ε_e yields $\frac{\partial k_o^*}{\partial \varepsilon_e} = \frac{hw(1-e^{-\gamma a_o^*}) + (1+\varepsilon_e+\varepsilon_o)hw\gamma e^{-\gamma a_o^*}\frac{\partial a_o^*}{\partial \varepsilon_e}}{2\beta} > 0$. Given that the profit function is $\pi_o^* = (p_{o,e}^* - c)(B_e - tw - hw_e e^{-\gamma a_o^*} - \beta p_{o,e}^* + \varepsilon_e hw(1-e^{-\gamma a_o^*})) + (p_{o,o}^* - c)(B_o - tw - hw_o - \beta p_{o,o}^* - \varepsilon_o hw(1-e^{-\gamma a_o^*})) - \frac{1}{2}\tau a_o^2$, differentiating with respect to ε_e using the envelope theorem yields $\frac{\partial \pi_o^*}{\partial \varepsilon_e} = (p_{o,e}^* - c)hw(1-e^{-\gamma a_o^*}) > 0$.

(ii) Differentiating F with respect to ε_o yields $\frac{\partial F}{\partial \varepsilon_o} = -(B_o - tw - hw - 2\varepsilon_o hw(1-e^{-\gamma a}) - \beta c)hw\gamma > 0$. Therefore, it is certain that $\frac{\partial a_o^*}{\partial \varepsilon_o} < 0$. Knowing that $p_{o,e}^* = \frac{B_e - tw + \varepsilon_e hw - (1+\varepsilon_e)hwe^{-\gamma a_o^*} + \beta c}{2\beta}$, differentiating with respect to ε_o yields $\frac{\partial p_{o,e}^*}{\partial \varepsilon_o} = ((1+\varepsilon_e)hw\gamma e^{-\gamma a_o^*}\frac{\partial a_o^*}{\partial \varepsilon_o})/2\beta < 0$; and knowing $p_{o,o}^* = \frac{B_o - tw - hw - \varepsilon_o hw(1-e^{-\gamma a_o^*}) + \beta c}{2\beta}$, differentiating with respect to ε_o yields $\frac{\partial p_{o,o}^*}{\partial \varepsilon_o} = (-hw(1-e^{-\gamma a_o^*}) - \varepsilon_o hw\gamma e^{-\gamma a_o^*}\frac{\partial a_o^*}{\partial \varepsilon_o})/2\beta$. Thus, it can be concluded that $\lim_{\varepsilon_o \to 0}\frac{\partial p_{o,o}^*}{\partial \varepsilon_o} < 0$ and $\lim_{\varepsilon_o \to +\infty}\frac{\partial p_{o,o}^*}{\partial \varepsilon_o} > 0$. Differentiating k_o^* with respect to ε_o yields $\frac{\partial k_o^*}{\partial \varepsilon_o} = \frac{hw(1-e^{-\gamma a_o^*}) + (1+\varepsilon_e+\varepsilon_o)hw\gamma e^{-\gamma a_o^*}\frac{\partial a_o^*}{\partial \varepsilon_o}}{2\beta}$, and since ε_e and ε_o are both greater than 0, it follows that $\frac{\partial k_o^*}{\partial \varepsilon_o} < 0$. Given that the profit function is $\pi_o^* = (p_{o,e}^* - c)(B_e - tw - hw_e e^{-\gamma a_o^*} - \beta p_{o,e}^* + \varepsilon_e hw(1-e^{-\gamma a_o^*})) + (p_{o,o}^* - c)(B_o - tw - hw_o - \beta p_{o,o}^* - \varepsilon_o hw(1-e^{-\gamma a_o^*})) - \frac{1}{2}\tau a_o^2$, differentiating with respect to ε_o using the envelope theorem yields $\frac{\partial \pi_o^*}{\partial \varepsilon_o} = -(p_{o,o}^* - c)hw(1-e^{-\gamma a_o^*}) < 0$.

B.6. Proof of Corollary 3

Given the observable case, the EWS level a_o^* satisfies

$$\begin{aligned}&\left(B_e - tw - hwe^{-\gamma a_o^*} + \varepsilon_e hw\left(1 - e^{-\gamma a_o^*}\right) - \beta c\right)(1+\varepsilon_e)\, hw\gamma\\&\quad - \left(B_o - tw - hw - \varepsilon_o hw\left(1 - e^{-\gamma a_o^*}\right) - \beta c\right)\varepsilon_o hw\gamma - 2\beta\tau a_o^* e^{\delta a_o^*} = 0.\end{aligned} \tag{B.4}$$

Substituting $\varepsilon_e = (1+\eta)\varepsilon_o$ into the above equation yields

$$\begin{aligned}&\left(B_e - tw - hwe^{-\gamma a} + (1+\eta)\varepsilon_o hw\left(1 - e^{-\gamma a}\right) - \beta c\right)(1+(1+\eta)\varepsilon_o)\, hw\gamma\\&\quad - \left(B_o - tw - hw - \varepsilon_o hw\left(1 - e^{-\gamma a}\right) - \beta c\right)\varepsilon_o hw\gamma - 2\beta\tau a e^{\delta a} = 0.\end{aligned} \tag{B.5}$$

Thus, F can be expressed as

$$\begin{aligned}F = &\left(B_e - tw - hwe^{-\gamma a} + (1+\eta)\varepsilon_o hw\left(1 - e^{-\gamma a}\right) - \beta c\right)(1+(1+\eta)\varepsilon_o)\, hw\gamma\\& - \left(B_o - tw - hw - \varepsilon_o hw\left(1 - e^{-\gamma a}\right) - \beta c\right)\varepsilon_o hw\gamma - 2\beta\tau a e^{\delta a}.\end{aligned} \tag{B.6}$$

The derivative of F with respect to ε_o is given by

$$\begin{aligned}\frac{\partial F}{\partial \varepsilon_o} &= hw\gamma\left[(1+\eta)\left(B_e - tw - hwe^{-\gamma a} + (1+\eta)\varepsilon_o hw\left(1 - e^{-\gamma a}\right) - \beta c\right)\right.\\&\quad\left. + (1+\eta)(1+(1+\eta)\varepsilon_o)\, hw\gamma\left(1 - e^{-\gamma a}\right)\right] - hw\gamma\left(B_o - tw - hw - 2\varepsilon_o hw\left(1 - e^{-\gamma a}\right) - \beta c\right).\end{aligned} \tag{B.7}$$

When $\frac{\partial F}{\partial \varepsilon_o} = 0$, namely $a_o^* = a_u^*$, the indifference point η_a^P satisfies $\frac{\partial F}{\partial \varepsilon_o} = 0$. The second derivative of $\frac{\partial F}{\partial \varepsilon_o}$ with respect to η is given by

$$\frac{\partial^2 F}{\partial \varepsilon_o \partial \eta} = hw\gamma\left[\left(B_e - tw - hwe^{-\gamma a} + 2(1+\eta)\varepsilon_o hw(1 - e^{-\gamma a}) - \beta c\right) + (1 + 2(1+\eta)\varepsilon_o)\, hw\left(1 - e^{-\gamma a}\right)\right] > 0.$$

In the observable case, the prices are given by

$$p_{o,e}^* = \frac{B_e - tw - hwe^{-\gamma a_o^*} + \varepsilon_e hw(1 - e^{-\gamma a_o^*}) + \beta c}{2\beta}$$

and

$$p_{o,o}^* = \frac{B_o - tw - hw - \varepsilon_o hw(1 - e^{-\gamma a_o^*}) + \beta c}{2\beta}.$$

The profit function is represented as follows:

$$\begin{aligned}\pi_o^* = &\left(p_{o,e}^* - c\right)\left(B_e - tw - hw_e e^{-\gamma a_o^*} - \beta p_{o,e}^* + (1+\eta)\varepsilon_o hw\left(1 - e^{-\gamma a_o^*}\right)\right)\\& + \left(p_{o,o}^* - c\right)\left(B_o - tw - hw_o - \beta p_{o,o}^* - \varepsilon_o hw\left(1 - e^{-\gamma a_o^*}\right)\right) - \frac{1}{2}\tau a_o^{*2}.\end{aligned} \tag{B.8}$$

The derivative of the profit function with respect to ε_o is given by

$$\frac{\partial \pi_o^*}{\partial \varepsilon_o} = \left(p_{o,e}^* - c\right)(1+\eta)hw\left(1 - e^{-\gamma a_o^*}\right) - \left(p_{o,o}^* - c\right)hw\left(1 - e^{-\gamma a_o^*}\right),$$

thus, when $\eta_\pi^P = \frac{p_{o,o}^* - c}{p_{o,e}^* - c} - 1$, it follows that $\frac{\partial \pi_o^*}{\partial \varepsilon_o} = 0$. Furthermore, since $\frac{\partial F}{\partial \varepsilon_o}$ can be rewritten as

$$\frac{\partial F}{\partial \varepsilon_o} = hw\gamma(1+\eta)(1+(1+\eta)\varepsilon_o)\, hw\left(1 - e^{-\gamma a}\right) + \varepsilon_o hw\left(1 - e^{-\gamma a}\right) > 0. \tag{B.9}$$

Hence, it is known that $\frac{\partial F}{\partial \varepsilon_o} > 0$ when $\eta_\pi^P = \frac{p_{o,o}^* - c}{p_{o,e}^* - c} - 1$. Additionally, since $\frac{\partial^2 F}{\partial \varepsilon_o \partial \eta} > 0$, there exists $\eta_\pi^P > \eta_a^P > 0$.

When $\eta = \eta_a^P$, $\frac{\partial a_o^*}{\partial \varepsilon_o} = 0$, implying $\frac{\partial p_{o,e}^*}{\partial \varepsilon_o} = \frac{1+\eta}{2\beta} hw(1 - e^{-\gamma a}) > 0$. On the other hand, when $\eta = \eta_\pi^P$, $\frac{\partial F}{\partial \varepsilon_o} > 0$, and thus $\frac{\partial a_o^*}{\partial \varepsilon_o} > 0$, leading to $\frac{\partial p_{o,e}^*}{\partial \varepsilon_o} > 0$. This condition holds for $\eta > \eta_a^P$, ensuring $\frac{\partial a_o^*}{\partial \varepsilon_o} > 0$ and $\frac{\partial p_{o,e}^*}{\partial \varepsilon_o} > 0$.

The derivative of $p_{o,o}$ with respect to ε_o is given by

$$\frac{\partial p_{o,o}}{\partial \varepsilon_o} = -\frac{hw\left(1 - e^{-\gamma a_o^*}\right) + \varepsilon_o hw\gamma e^{-\gamma a} \frac{\partial a_o^*}{\partial \varepsilon_o}}{2\beta},$$

which is negative when $\eta = \eta_o^a$ and $\frac{\partial a_o^*}{\partial \varepsilon_o} = 0$. The monotonicity of $p_{o,o}$ with respect to ε_o cannot be determined in other cases.

Similarly, the derivative of k_o^* with respect to ε_o is given by

$$\frac{\partial k_o^*}{\partial \varepsilon_o} = \frac{\gamma hwe^{-\gamma a_o^*} \frac{\partial a_o^*}{\partial \varepsilon_o} + (2+\eta)hw\left(1 - e^{-\gamma a_o^*}\right) + (2+\eta)\varepsilon_o hw\gamma e^{-\gamma a_o^*} \frac{\partial a_o^*}{\partial \varepsilon_o}}{2\beta},$$

and when $\eta = \eta_a^P$, $\frac{\partial a_o^*}{\partial \varepsilon_o} = 0$, resulting in $\frac{\partial k_o^*}{\partial \varepsilon_o} = \frac{2+\eta}{2\beta} hw(1 - e^{-\gamma a_o^*}) > 0$. Furthermore, when $\eta = \eta_\pi^P$, $\frac{\partial F}{\partial \varepsilon_o} > 0$, leading to $\frac{\partial a_o^*}{\partial \varepsilon_o} > 0$, and thus $\frac{\partial k_o^*}{\partial \varepsilon_o} > 0$. This condition holds for $\eta > \eta_a^P$, ensuring $\frac{\partial a_o^*}{\partial \varepsilon_o} > 0$ and $\frac{\partial k_o^*}{\partial \varepsilon_o} > 0$, indicating $k_o^* > k_u^*$.

B.7. Proof of Proposition 3

Given the profit function

$$\pi_u^D = (p_e - c)\left(B_e - tw_e - hw_e e^{-\gamma a} - \beta p_e\right) + (p_o - c)(B_o - tw_o - hw_o - \beta p_o) - \frac{c}{w_e} - \frac{c}{w_o} - \frac{1}{2}\tau a^2,$$

the partial derivatives with respect to p_e and p_o are

$$\frac{\partial \pi_u^D}{\partial p_e} = B_e - tw_e - hw_e e^{-\gamma a} - 2\beta p_e + \beta c$$

and

$$\frac{\partial \pi_u^D}{\partial p_o} = B_o - tw_e - hw_e - 2\beta p_o + \beta c.$$

Setting $\frac{\partial \pi_u^D}{\partial p_e} = 0$ and $\frac{\partial \pi_u^D}{\partial p_o} = 0$, we obtain the optimal prices $p_{u,e}^D = \frac{B_e - tw_e - hw_e e^{-\gamma a} + \beta c}{2\beta}$ and $p_{u,o}^D = \frac{B_o - tw_o - hw_o + \beta c}{2\beta}$.

Substituting these prices back into the profit function, we have

$$\pi_u^D = \frac{\left(B_e - tw_e - hw_e e^{-\gamma a} - \beta c\right)^2}{4\beta} + \frac{\left(B_o - tw_o - hw_0 - \beta c\right)^2}{4\beta} - \frac{c}{w_e} - \frac{c}{w_o} - \frac{1}{2}\tau a^2.$$

Taking the first and second derivatives of π_u^D with respect to a, we get

$$\frac{\partial \pi}{\partial a} = \frac{\left(B_e - tw_e - hw_e e^{-\gamma a} - \beta c\right) hw_e \gamma e^{-\gamma a}}{2\beta} - \tau a$$

and

$$\frac{\partial^2 \pi}{\partial a^2} = -\frac{\gamma(B_e - tw_e - 2hw_e e^{-\gamma a} - \beta c)hw_e \gamma^2 e^{-\gamma a}}{2\beta} - \tau.$$

Therefore, when $B_e - tw_e - 2hw_e - \beta c \geq 0$, there exists an optimal EWS level a_u^D satisfying $(B_e - tw_e - hw_e e^{-\gamma a_u^D} - \beta c)hw_e\gamma - 2\beta\tau a_u^D e^{\gamma a_u^D} = 0$.

Hence, the optimal prices and profit function are represented as follows:

$$p_{u,e}^D = \frac{B_e - tw_e - hw_e e^{-\gamma a_u^D} + \beta c}{2\beta}, \quad p_{u,o}^D = \frac{B_o - tw_o - hw_o + \beta c}{2\beta},$$

and

$$\pi_u^D = \left(p_{u,e}^D - c\right)\left(B_e - tw_e - hw_e e^{-\gamma a_u^D} - \beta p_{u,e}^D\right) + \left(p_{u,o}^D - c\right)\left(B_o - tw_o - hw_o - \beta p_{u,o}^D\right) - \frac{c}{w_e} - \frac{c}{w_o} - \frac{1}{2}\tau a_u^D.$$

B.8. Proof of Corollary A.3

The EWS level a_u^D satisfies $(B_e - tw_e - hw_e e^{-\gamma a_u^D} - \beta c)hw_e\gamma - 2\beta\tau a_u^D e^{\gamma a_u^D} = 0$. Let $G_1 = (B_e - tw_e - hw_e e^{-\gamma a_u^D} - \beta c)hw_e\gamma - 2\beta\tau a_u^D e^{\gamma a_u^D}$. Therefore, $\frac{\partial G_1}{\partial w_e} = (B_e - tw_e - 2hw_e e^{-\gamma a} - \beta c)h\gamma > 0$, and hence, there must exist $\frac{\partial a_u^D}{\partial w_e} = -\frac{\frac{\partial G_1}{\partial w_e}}{\frac{\partial G_1}{\partial a}} > 0$. However, $\frac{\partial G_1}{\partial w_o} = 0$, so $\frac{\partial a_u^D}{\partial w_o} = -\frac{\frac{\partial G_1}{\partial w_o}}{\frac{\partial G_1}{\partial a}} = 0$.

The profit function is given by $\pi_u^D = (p_{u,e}^D - c)(B_e - tw_e - hw_e e^{-\gamma a_u^D} - \beta p_{u,e}^D) + (p_{u,o}^D - c)(B_o - tw_o - hw_o - \beta p_{u,o}^D) - \frac{c}{w_e} - \frac{c}{w_o} - \frac{1}{2}\tau a_u^D$. According to the envelope theorem, $\frac{\partial \pi_u^D}{\partial w_e} = -(p_{u,e}^D - c)(t + he^{-\gamma a_u^D}) + \frac{c}{{w_e}^2}$ and $\frac{\partial \pi_u^D}{\partial w_o} = -(p_{u,o}^D - c)(t+h) + \frac{c}{{w_o}^2} < 0$. When $(p_{u,e}^D - c)(t + he^{-\gamma a_u^D}) < \frac{c}{{w_e}^2}$, $\frac{\partial \pi_u^D}{\partial w_e} > 0$; and when $(p_{u,e}^D - c)(t + he^{-\gamma a_u^D}) > \frac{c}{{w_e}^2}$, $\frac{\partial \pi_u^D}{\partial w_e} < 0$. Similarly, when $(p_{u,o}^D - c)(t+h) < \frac{c}{{w_o}^2}$, $\frac{\partial \pi_u^D}{\partial w_o} > 0$; and when $(p_{u,o}^D - c)(t+h) > \frac{c}{{w_o}^2}$, $\frac{\partial \pi_u^D}{\partial w_o} < 0$.

B.9. Proof of Proposition 4

Given the profit function $\pi = (p_e - c)(B_e - tw_e - hw_e e^{-\gamma a} + \varepsilon_e[t(w - w_e) + h(w - w_e e^{-\gamma a})] - \beta p_e) + (p_o - c)(B_o - tw - hw - \varepsilon_o[t(w - w_e) + h(w - w_e e^{-\gamma a})] - \beta p_o) - \frac{c}{w_e} - \frac{c}{w} - \frac{1}{2}\tau a^2$, where $\frac{\partial \pi}{\partial p_e} = B_e - tw_e - hw_e e^{-\gamma a} + \varepsilon_e[t(w - w_e) + h(w - w_e e^{-\gamma a})] - 2\beta p_e$ and $\frac{\partial \pi}{\partial p_o} = B_o - tw - hw - \varepsilon_o[t(w - w_e) + h(w - w_e e^{-\gamma a})] - 2\beta p_o$. Setting $\frac{\partial \pi}{\partial p_e} = 0$ and $\frac{\partial \pi}{\partial p_o} = 0$, we obtain $p_{o,e}^D = \frac{B_e - tw_e - hw_e e^{-\gamma a} + \varepsilon_e[t(w - w_e) + h(w - w_e e^{-\gamma a})] + \beta c}{2\beta}$ and $p_{o,o}^D = \frac{B_o - tw - hw - \varepsilon_o[t(w - w_e) + h(w - w_e e^{-\gamma a})] + \beta c}{2\beta}$. Substituting these prices into the profit function, we get $\pi = \frac{(B_e - tw_e - hw_e e^{-\gamma a} + \varepsilon_e[t(w - w_e) + h(w - w_e e^{-\gamma a})] - \beta c)^2}{4\beta} + \frac{(B_o - tw - hw - \varepsilon_o[t(w - w_e) + h(w - w_e e^{-\gamma a})] + \beta c)^2}{4\beta} - \frac{c}{w_e} - \frac{c}{w} - \frac{1}{2}\tau a^2$. The derivative of π with respect to the EWS level a is $\frac{\partial \pi}{\partial a} = \frac{[B_e - tw_e - hw_e e^{-\gamma a} + \varepsilon_e[t(w - w_e) + h(w - w_e e^{-\gamma a})] - \beta c](1 + \varepsilon_e)hw_e\gamma e^{-\gamma a}}{2\beta} - \frac{[B_o - tw - hw - \varepsilon_o[t(w - w_e) + h(w - w_e e^{-\gamma a})] + \beta c]\varepsilon_o hw_e\gamma e^{-\gamma a}}{2\beta} - \tau a$, and the second derivative is

$$\begin{aligned}\frac{\partial^2 \pi}{\partial a^2} = &\frac{-\left[B_e - tw_e - 2\left(1 + \varepsilon_e\right)hw_e e^{-\gamma a} + \varepsilon_e\left[t\left(w - w_e\right) + w\right] - \beta c\right]\left(1 + \varepsilon_e\right)hw_e\gamma e^{-\gamma a}}{2\beta} \\ &+ \frac{\left[B_o - tw - hw + 2\varepsilon_o hw_e e^{-\gamma a} - \varepsilon_o\left[t\left(w - w_e\right) + hw\right] - \beta c\right]\varepsilon_o hw_e\gamma e^{-\gamma a}}{2\beta} - \tau.\end{aligned} \tag{B.10}$$

Thus, when $\frac{B_e - tw_e - 2(1+\varepsilon_e)hw_e + \varepsilon_e[t(w - w_e) + w] - \beta c}{B_o - tw - hw + 2\varepsilon_o hw_e - \varepsilon_o[t(w - w_e) + hw] - \beta c} \geq \frac{\varepsilon_o}{1 + \varepsilon_e}$, there exists an optimal a_o^D satisfying $[B_e - tw_e - hw_e e^{-\gamma a_o^D} + \varepsilon_e[t(w - w_e) + h(w - w_e e^{-\gamma a_o^D})] - \beta c](1 + \varepsilon_e)hw_e\gamma - [B_o - tw - hw - \varepsilon_o[t(w - w_e) + h(w - w_e e^{-\gamma a_o^D})] + \beta c]\varepsilon_o hw_e\gamma - 2\beta\tau a e^{\gamma a_o^D} = 0$. In this case, the optimal prices and profit function are given by $p_{o,e}^D = \frac{B_e - tw_e - hw_e e^{-\gamma a_o^D} + \varepsilon_e[t(w - w_e) + h(w - w_e e^{-\gamma a_o^D})] + \beta c}{2\beta}$, $p_{o,o}^D = \frac{B_o - tw - hw - \varepsilon_o[t(w - w_e) + h(w - w_e e^{-\gamma a_o^D})] + \beta c}{2\beta}$, and $\pi_o^D = (p_{o,e}^D - c)\lambda_e + (p_{o,o}^D - c)\lambda_o - \frac{c}{w_e} - \frac{c}{w} - \frac{1}{2}\tau a_o^D$.

B.10. Proof of Corollary A.4

(i) Because $\varepsilon_e = (1+\eta)\varepsilon_o$, we can obtain

$$G_2 = \left[B_e - \left(t + he^{-\gamma a_o^D}\right)(w-\theta) + (1+\eta)\varepsilon_o\left[t\theta + h\left(w - (w-\theta)e^{-\gamma a_o^D}\right)\right] - \beta c\right] \times \left(1 + (1+\eta)\varepsilon_o\right)h(w-\theta)\gamma - \left[B_o - tw - hw - \varepsilon_o\left[t\theta + h\left(w-(w-\theta)e^{-\gamma a_o^D}\right)\right] + \beta c\right]\varepsilon_o h(w-\theta)\gamma - 2\beta\tau a e^{\gamma a_o^D} = 0. \tag{B.11}$$

The derivative of G_2 with respect to θ yields the following function:

$$\frac{\partial G_2}{\partial\theta} = -\left(1+(1+\eta)\varepsilon_o\right)h\gamma\left[B_e - \left(t + he^{-\gamma a_o^D}\right)(w-\theta) + (1+\eta)\varepsilon_o\left[t\theta + h\left(w-(w-\theta)e^{-\gamma a_o^D}\right)\right] - \beta c\right] + \left(1+(1+\eta)\varepsilon_o\right)h(w-\theta)\gamma\left[t + he^{-\gamma a_o^D} + (1+\eta)\varepsilon_o\left(t + he^{-\gamma a_o^D}\right)\right] + \left[B_o - tw - hw - \varepsilon_o\left[t\theta + h\left(w-(w-\theta)e^{-\gamma a_o^D}\right)\right] + \beta c\right]\varepsilon_o h\gamma + \varepsilon_o h(w-\theta)\gamma\varepsilon_o\left(t + he^{-\gamma a_o^D}\right). \tag{B.12}$$

Thus we know

$$\frac{\partial G_2}{\partial\theta} < -\left(1+\varepsilon_e\right)h\gamma\left[B_e - tw_e - 2\left(1+\varepsilon_e\right)hw_e + \varepsilon_e\left[t\left(w-w_e\right) + hw\right] - \beta c - \left(1+\varepsilon_e\right)tw_e\right] + \left[B_o - tw - hw + 2\varepsilon_o hw_e - \varepsilon_o\left[t\left(w-w_e\right)+hw\right] - \beta c + \varepsilon_o tw_e\right]\varepsilon_o h\gamma. \tag{B.13}$$

Let $\bar{H}_e$ be defined as

$$\bar{H}_e = \frac{B_e - tw_e - 2(1+\varepsilon_e)hw_e + \varepsilon_e[t(w-w_e)+hw] - \beta c - (1+\varepsilon_e)tw_e}{B_o - tw - hw + 2\varepsilon_o hw_e - \varepsilon_o[t(w-w_e)+hw] - \beta c + \varepsilon_o tw_e},$$

then, we can determine that when $\bar{H}_e \geq \frac{\varepsilon_o}{1+\varepsilon_e}$, we have $\frac{\partial a_o^D}{\partial\theta} < 0$. Once $\bar{H}_e \geq \frac{\varepsilon_o}{1+\varepsilon_e} > \bar{H}_e$, we can have $\frac{\partial a_o^D}{\partial\theta} > 0$.

(ii) Additionally, we know that

$$\pi_o^D = \left(p_{o,e}^D - c\right)\left[B_e - t(w-\theta) - h(w-\theta)e^{-\gamma a} + \varepsilon_e\left[t\theta + h\left(w-(w-\theta)e^{-\gamma a}\right)\right] - \beta p_e\right] + \left(p_{o,o}^D - c\right)\left[B_o - tw - hw - \varepsilon_o\left[t\theta + h\left(w-(w-\theta)e^{-\gamma a}\right)\right] - \beta p_o\right] - \frac{c}{(w-\theta)} - \frac{c}{w} - \frac{1}{2}\tau {a_o^D}^2. \tag{B.14}$$

According to the envelope theorem, the derivative of π_o^D with respect to θ is given by

$$\frac{\partial\pi_o^D}{\partial\theta} = (p_{o,e}^D - c)[t + he^{-\gamma a} + \varepsilon_e t + \varepsilon_e he^{-\gamma a}] - (p_{o,o}^D - c)[\varepsilon_o t + \varepsilon_o he^{-\gamma a}] - \frac{c}{(w-\theta)^2}.$$

The difference in profits is related to the unit service cost c. We need to determine the threshold for c, and we also know that $p_{o,e}^D > p_{o,o}^D$. When $\lim_{c\to 0}(\partial\pi_o^D/\partial\theta) = p_{o,e}^D[t+he^{-\gamma a}+\varepsilon_e t+\varepsilon_e he^{-\gamma a}] - p_{o,o}^D[\varepsilon_o t+\varepsilon_o he^{-\gamma a}] > 0$, and when $\lim_{c\to p_{o,e}^D}(\partial\pi_o^D/\partial\theta) = -c/(w-\theta)^2 < 0$, there must exist $\hat{c}$ such that

$$\left(p_{o,e}^D - \hat{c}\right)\left[t + he^{-\gamma a} + \varepsilon_e t + \varepsilon_e he^{-\gamma a}\right] = \left(p_{o,o}^D - \hat{c}\right)\left[\varepsilon_o t + \varepsilon_o he^{-\gamma a}\right] + \frac{\hat{c}}{(w-\theta)^2},$$

when $c \leq \hat{c}$, $(\partial\pi_o^D/\partial\theta) \geq 0$ exists. In this case, $\pi_o^D \geq \pi_o^*$. When $c > \hat{c}$, $(\partial\pi_o^D/\partial\theta) < 0$ exists, and we have $\pi_o^D < \pi_o^*$.

B.11. Proof of Corollary 4

Similarly to the proof of Corollary 3.

B.12. Proof of Corollaries 5 and 6

We already know that under dedicated capacity, when EWS is unobservable, the EWS level a_u^D satisfies $(B_e - tw_e - hw_e e^{-\gamma a_u^D} - \beta c)hw_e\gamma - 2\beta\tau a e^{\gamma a_u^D} = 0$. Let $G = (B_e - tw_e - hw_e e^{-\gamma a_u^D} - \beta c)hw_e\gamma - 2\beta\tau a e^{\gamma a_u^D}$. Taking the derivative of G with respect to w_e, we get $\partial G/\partial w_e = (B_e - tw_e - 2hw_e e^{-\gamma a_u^D} - \beta c)h\gamma > 0$. Therefore, we conclude that $\partial a_u^D/\partial w_e > 0$, implying that there must exist $a_u^D < a_u^*$.

We know the profit function is given by $\pi_u^D = (p_{u,e}^D - c)(B_e - tw_e - hw_e e^{-\gamma a_u^D} - \beta p_{u,e}^D) + (p_{u,o}^D - c)(B_o - tw_o - hw_o - \beta p_{u,o}^D) - c/w_e - c/w_o - 1/2\tau a_u^{D2}$. Differentiating π_u^D with respect to w_e, we obtain $\partial \pi_u^D/\partial w_e = -(p_{u,e}^D - c)he^{-\gamma a_u^D} + c/{w_e}^2$. We need to determine the threshold for c. When $\lim_{c\to 0}(\partial\pi_u^D/\partial w_e) = -p_{u,e}^D he^{-\gamma a_u^D} < 0$, and $\lim_{c\to p_{u,e}^D}(\partial\pi_u^D/\partial w_e) = c/{w_e}^2 > 0$. Therefore, there must exist $\bar{c}$ such that $he^{-\gamma a_u^D} = \bar{c}/((p_{u,e}^D - \bar{c}){w_e}^2)$. When $c \le \bar{c}$, $(\partial\pi_u^D/\partial w_e) \le 0$, implying $\pi_u^D \ge \pi_u^*$. However, when $c > \bar{c}$, $(\partial\pi_u^D/\partial w_e) > 0$, and thus, we have $\pi_u^D < \pi_u^*$.

B.13. Proof of Corollaries 7 and 8

Can be directly derived from Corollary A.4.

Acknowledgements

This study is supported by the Major Project of National Social Science Foundation of China (Grant No. 17ZDA030) and the National Natural Science Foundation of China (Grant No. 71831005, 71502044).

References

[1] C.S. Airlines, Airport VIP lounge display (2023). http://www.csair.cn/cn/tourguide/airport_service/priority_pass/domestic/18h9mvejd5308.shtml.
[2] Z. Aksin, B. Ata, S. Emadi and C. Su, Impact of delay announcements in call centers: an empirical approach. *Oper. Res.* **65** (2017) 242–265.
[3] M. Armony, G. Roels and H. Song, Pooling queues with strategic servers: the effects of customer ownership. *Oper. Res.* **69** (2021) 13–29.
[4] E.J. Arnould, L.L. Price and P. Tierney, Communicative staging of the wilderness servicescape. *Serv. Ind. J.* **18** (1998) 90–115.
[5] A. Bassamboo and R. Ibrahim, A general framework to compare announcement accuracy: static vs. LES-based announcement. *Manag. Sci.* **67** (2021) 4191–4208.
[6] S. Benjaafar, Performance bounds for the effectiveness of pooling in multi-processing systems. *Eur. J. Oper. Res.* **87** (1995) 375–388.
[7] P. Cao, S. He, J. Huang and Y. Liu, To pool or not to pool: queueing design for large-scale service systems. *Oper. Res.* **69** (2021) 1866–1885.
[8] M. Chen, D. Zhao, Y. Gong and Z. Hong, Reference-dependent preferences in the on-demand service newsvendor with self-scheduling capacity. *Int. J. Prod. Econ.* **234** (2021) 108043.
[9] A. Chernev, When more is less and less is more: the role of ideal point availability and assortment in consumer choice. *J. Consum. Res.* **30** (2003) 170–183.
[10] C. Deng, Guide for priority passenger traveling by train (2023). https://www.chinatrainbooking.com/news/guide-for-passengers.html.
[11] P. Feldman, A.E. Frazelle and R. Swinney, Managing relationships between restaurants and food delivery platforms: conflict, contracts, and coordination. *Manag. Sci.* **69** (2023) 812–823.
[12] N. Gans, G. Koole and A. Mandelbaum, Telephone call centers: tutorial, review, and research prospects. *Manuf. Serv. Oper. Manag.* **5** (2003) 79–141.
[13] F. Gao and X. Su, Omnichannel service operations with online and offline self-order technologies. *Manag. Sci.* **64** (2018) 3595–3608.
[14] S.M. Gilbert and Z.K. Weng, Incentive effects favor nonconsolidating queues in a service system: the principalagent perspective. *Manag. Sci.* **44** (1998) 1662–1669.
[15] O. Jouini, O. Aksin, F. Karaesmen, M. Aguir and Y. Dallery, Call center delay announcement using a newsvendor-like performance criterion. *Prod. Oper. Manag.* **24** (2015) 587–604.

[16] M. Kostecki, Waiting lines as a marketing issue. *Eur. Manag. J.* **14** (1996) 295–303.
[17] P. Kotler, Atmospherics as a marketing tool. *J. Retail.* **49** (1974) 48–64.
[18] Y. Kuo and C. Wu, Satisfaction and post-purchase intentions with service recovery of online shopping websites: perspectives on perceived justice and emotions. *Int. J. Inf. Manag.* **32** (2009) 127–138.
[19] S.M. Leong, S.H. Ang and L.H.L. Low, Effects of physical environment and locus of control on service evaluation: a replication and extension. *J. Retail. Consum. Serv.* **4** (1997) 231–237.
[20] X. Li, Z. Lian and Y. Shi, Capacity co-opetition in service clusters with waiting-area entertainment. *SSRN Electron. J.* (2021).
[21] X. Liu and P.T. Popkowski Leszczyc, The reference price effect of historical price lists in online auctions. *J. Retail. Consum. Serv.* **71** (2023) 103183.
[22] J. Liu, S. Zhang, H. Zhao and S. Liu, Research on service mechanism and pricing based on customer fairness preference. *J. Econ. Manag. Strategy* **27** (2019) 175–184.
[23] D. Maister, The psychology of waiting lines, in The Service Encounter, edited by J.A. Czepiel, M. Solomon and C.S. Surprenant. Lexington Books, Lexington (1985).
[24] S.K. Mukhopadhyay, X. Su and S. Ghose, Motivating retail marketing effort: optimal contract design. *Prod. Oper. Manag.* **18** (2009) 197–211.
[25] N. NetEase, Bank cancels seats for vip customers only? (2020). https://m.weibo.cn/1974561081/4544305836856348.
[26] A. Rafaeli, G. Barron and K. Haber, The effects of queue structure on attitudes. *J. Serv. Res.* **5** (2002) 125–139.
[27] K.N. Rajendran and G.J. Tellis, Contextual and temporal components of reference price. *J. Mark.* **58** (1994) 22–34.
[28] M. Rubinovitch, The slow server problem: a queue with stalling. *J. Appl. Probab.* **22** (1985) 879–892.
[29] L. Sams, Inside the secret world of ultra luxury VIPs (2023). https://https://www.afr.com/life-and-luxury/fashion-and-style/inside-the-secret-world-of-ultra-luxury-vips-20230123-p5cerb.
[30] H.L. Shi, S.Q. Chen, Y.M. Wang and Y. Huang, Cross-efficiency evaluation method taking management objectives as reference points from peer perspective. *Asia-Pac. J. Oper. Res.* **39** (2022) 2150040.
[31] S. Singh, M. Delasay and A. Scheller-Wolf, Real-time delay announcement under competition. *Prod. Oper. Manag. 32* (2023) 863–881.
[32] D.R. Smith and W. Whitt, Resource sharing for efficiency in traffic systems. *Bell Syst. Tech. J.* **60** (1981) 39–55.
[33] N. Sohu, What equivalent value-added services make customers willing to queue up? (2017). https://www.sohu.com/a/167435621_672964.
[34] N. Sohu, Airport lounges provide high quality services for easy travel. What are their services? (2023). http://news.sohu.com/a/666674612_120897976.
[35] H. Song, A.L. Tucker and K.L. Murrell, The diseconomies of queue pooling: an empirical investigation of emergency department length of stay. *Manag. Sci.* **61** (2015) 3032–3053.
[36] K. Sun, Y. Liu and X. Li, Make waiting not to seem like waiting: capacity management of waiting-area entertainment. *Oper. Res. Lett.* **50** (2022) 745–752.
[37] N. Sunar, Y. Tu and S. Ziya, Pooled vs. dedicated queues when customers are delay-sensitive. *Manag. Sci.* **67** (2021) 3785–3802.
[38] A. Terzi, K. Koedijk, C. Noussair and R. Pownall, Reference point heterogeneity. *Front. Psychol.* **7** (2016) 1–10.
[39] K.A. Tversky, Prospect theory: an analysis of decision under risk. *Econometrica* **47** (1979) 263–291.
[40] A. Tversky and D. Kahneman, Loss aversion in riskless choice: a reference-dependent model. *Q. J. Econ.* **106** (1991) 1039–1061.
[41] Z. Wang, L. Yang, S. Cui, S. Ülkü and Y.-P. Zhou, Pooling agents for customer-intensive services. *Oper. Res.* **71** (2023) 860–875.
[42] L. Yang, P. Guo and Y. Wang, Service pricing with loss-averse customers. *Oper. Res.* **66** (2018) 761–777.
[43] J.F. Yates and E.R. Stone, The Risk Construct. John Wiley & Sons (1992) 1–25.
[44] X. Yuan and H.B. Hwarng, Social interactions in service operations: demand-inducing vs. congestion-aggravating. *J. Oper. Res. Soc.* (2023) 1–23.
[45] X. Yuan, T. Dai, L.G. Chen and S. Gavirneni, Co-opetition in service clusters with waiting-area entertainment. *Manuf. Serv. Oper. Manag.* **23** (2021) 106–122.
[46] W. Zhan, M. Jiang and C. Li, The service strategy of customer-intensive services under advertising effects. *Kybernetes* **50** (2021) 2453–2470.
[47] W. Zhan, M. Jiang and X. Wang, The optimal capacity decision of the catering merchant in omnichannel – service, production and delivery capacity. *Kybernetes* **53** (2024) 1958–1986.

[48] J.C. Zhang, C. Gang and Z.M. Liu, Strategic behavior in a vacation queue with delayed observations. *RAIRO-Oper. Res.* **55** (2021) 2423–2437.

[49] J. Zhou, Reference dependence and market competition. *J. Econ. Manag. Strategy* **20** (2011) 1073–1097.

[50] L. Zhou, QoE-driven delay announcement for cloud mobile media. *IEEE Trans. Circuits Syst. Video Technol.* **27** (2017) 84–94.

[51] W. Zhou, D. Wang, W. Huang and P. Guo, To pool or not to pool? The effect of loss aversion on queue configurations. *Prod. Oper. Manag.* **30** (2021) 4258–4272.