Ph. D. Dissertation in Engineering

Semiparametric Modeling of Consumer Behavior Incorporating Time-Varying Effects

- Focusing on Updating Expectations and Perceptions in Platform Service Use-

August 2022

Graduate School of Seoul National University

Technology Management, Economics, and Policy Program

Justine Jihyun Kim

Semiparametric Modeling of Consumer Behavior Incorporating Time-Varying Effects

- Focusing on Updating Expectations and Perceptions in Platform Service Use -

지도교수 이종수

이 논문을 공학박사학위 논문으로 제출함
2022 년 7 월

서울대학교 대학원

협동과정 기술경영경제정책 전공

김지현

김지현의 공학박사학위 논문을 인준함
2022 년 6 월

위 원 장 이 정 동 (인)

부위원장 이 종 수 (인)

위 원 조 영 상 (인)

위 원 구 윤 모 (인)

위 원 신 정 우 (인)

Abstract

Semiparametric Modeling of Consumer Behavior Incorporating Time-Varying Effects

- Focusing on Updating Expectations and Perceptions in Platform Service Use -

Justine Jihyun Kim
Technology Management, Economics, and Policy Program
The Graduate School
Seoul National University

The marketing literature clearly defines that consumers' intention to repurchase a product or continue to use a service depends primarily on their prior experience of using them, and that continued user satisfaction is considered the key to building and retaining a loyal base of long-term customers. However, most existing studies use static utility models to explain consumer behavior in platform services and therefore do not adequately reflect the time-varying effects of continued use of the service. In addition, cross-sectional studies of consumers' continued use of services cannot provide an accurate view of how customers' expectations and perceptions of the product/service may change over time. Therefore, dynamic longitudinal studies are needed to determine how customers update their expectations and perceptions through experience and how this may affect customer satisfaction and/or behavior. This study aims to fill this gap by employing a dynamic utility model to explain consumer behavior in a platform economy where services are used repeatedly. Through an empirical study, we examine the time-varying effects of covariates in explaining consumers' use of ride-hailing platforms by first identifying the effect of updating expectations and perceptions with repeated use, thereby extending upon the expectation-confirmation theory. In the second part of this study, we observe the temporal effects on consumers' usage behavior through semiparametric modeling. The results of this study are expected to add to the literature on consumer behavior by presenting how the discrepancy between updated service expectations and actual service delivery, as well as updated perceptions, affect consumer behavior in platform services and by demonstrating seasonality in services with repeated use.

Keywords: Consumer Behavior, Repeated Use, Semi-parametric Modeling, Expectations Updating, Time-varying Effects, Platform Services

Student Number: 2017-32883

Contents

List of Tables

List of Figures

Chapter 1. Introduction

1.1 Research Background

In the sharing economy, an online platform that brings together supply and demand is a driving force for growth, innovation and competition. They enable a business model that creates value by connecting service providers and consumers, enabling businesses and consumers to take advantage of opportunities in the digital economy (Parker et al., 2016). The mobile platform is a representative example of information technology (IT), where consumers can access the necessary information and services without time and space constraints through the mobile data service provided. It provides a service based consumer's location and can promptly provide the necessary information they need. Today, the platform revolution is no longer limited to the retail or high-tech industries, but can be observed in several industries (Parker et al., 2010).

A representative business strategy of platform service providers is consumer lock-in, where consumers are dependent on a single producer or provider for a particular service and cannot switch to another provider without significant cost or inconvenience (Arthur, 1989; Farrel & Klemperer, 2007). In a platform environment, once consumers have chosen a predominantly used service, they continue to invest in using that particular product rather than replacing the product they are already using, creating a potential barrier to entry for potential competitors (Arthur, 1989; Barney, 1991). However, users also have the option of using multiple services, making it easy to find a substitute that offers a similar service, as the obvious difference between the various services is usually difficult to discern. Since users, once they choose a service, tend to use it repeatedly, the churn rate may increase if user satisfaction cannot be maintained over time.

Therefore, studies in the field of management information and marketing have attempted to explore the factors that induce consumers' continued use of various services. There is an agreement in the marketing literature that consumer satisfaction with the perceived quality (including pricing) of services is critical to customer retention, and that the value of a service depends on its quality and price (Steiner et al., 2013). Moreover, while there is a general consensus that a person's decision to reuse a service, as opposed to the decision to use it for the first time, depends on the expected value of the service – that is, the benefits that is derived from using the service relative to expectations. Therefore, individual evaluation of the value of a service is likely to depend on his or her previous experience with services (Ganesh et al, 2000; Kalwani and Narayandas, 1995; Bolton et al., 2006). Similarly, Helson, (1964) established that a person's perception of a new stimulus is formed relative to the reference value accumulated through experience through adaptation level theory. This suggests, for example, that observed prices and internal reference prices that are determined by previous information to which consumers have been exposed are compared (Blattberg and Neslin, 1990; Kneib et al., 2007). Thus, some studies have designed choice models to include additional covariates that reflect the discrepancy between observed and reference prices (Kalwani et al. 1990, or Kalyanaram and Little 1994, Kneib et al., 2007). Similarly, studies have shown that perceptions of service quality are positively correlated with customer retention (e.g., Boulding et al. 1993).

Furthermore, marketing literature states that a consumer's intention to purchase or use a product or service again is mainly determined by previous experience with that product or service (Anderson and Sullivan, 1993; Oliver, 1980, 1993, Steiner et al., 2013). Additionally, sustained user satisfaction considered key to building and maintaining a loyal base of long-term consumers (Bolton et al., 2006). However, existing studies on demand-side adoption behavior of platform services do not adequately reflect the time-varying effects of continued use of the service; they utilize static utility models to explain consumer behavior for these platform services. Moreover, cross-sectional studies of consumers' continued use of services are unable to capture an accurate view of how customers' expectations and perceptions of products/services may change as a function of their consumption experiences and the impact of these changes on subsequent cognitive processes. Therefore, a dynamic longitudinal study is needed to determine how customers update their expectations and perceptions through experience, and how these, in turn influence consumer satisfaction and/or behavior. This study aims to fill this gap by implementing a dynamic utility model to explain consumer behavior in the platform economy, where repeated use of services occurs.

Through an empirical study, we examine the time-varying effects of covariates for explaining consumers' use of ride-hailing platforms by, first, identifying the effect of updated expectations (UE) and updated perceptions (UP) with repeated use and then incorporating models based on penalized splines, a semiparametric approach. Second, we observe temporal effects in consumers' usage behavior through semiparametric modeling. The results of this study are expected to add to the literature on consumer behavior by presenting how the discrepancy between updated service expectations and actual service delivery, as well as updated perceptions, affect consumer behavior in platform services and by identifying seasonality in services with repeated use.

1.2 Research Objectives

This dissertation aims to incorporate time-varying effects of covariates in explaining consumers' use of platform service by using models based on P-splines, a semiparametric approach. In doing so, this study first incorporates accumulated experience effect rising from repeated use of service, elaborating on how the described and/or updated 'Service Gap' influence consumers' usage behavior of the platform. While it is assumed that the customer's expectation of the service is either (1) described (specified) by the service provider or (2) formed by their own past experiences with repeated use, the study is expected to describe how consumers' usage experience can update consumer expectations and perceptions, and they influence consumer's platform usage behavior over time. Thereafter, by incorporating time-varying effects of covariates, stream-of-time effects (seasonality) in consumer behaviors can also be observed. The empirical analysis will be conducted in the context of ride-hailing platforms. The study is a meaningful addition to the literature that aims to understand platform service adoption from the demand-side perspective and provides important implications that can be applied in practice when considering strategies to retain and improve the loyalty of ride-hailing platform users.

Chapter 2. Literature Review

2.1 Studies on Consumer Behavior

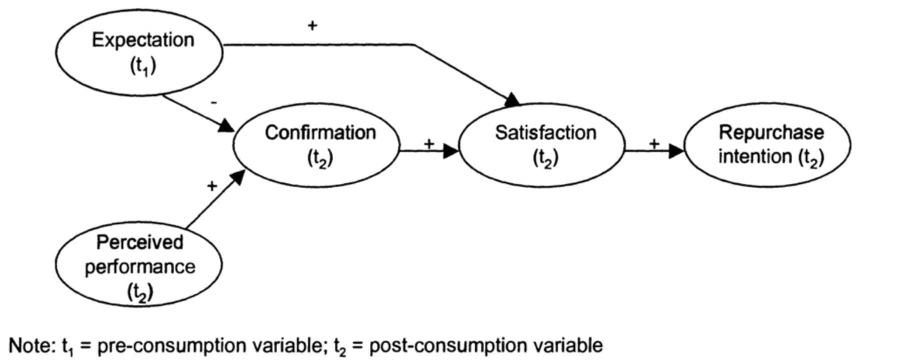
2.1.1 Expectation-Confirmation Theory

Expectation Confirmation Theory (ECT), which states that consumers' satisfaction with previous use of that product or service primarily determines their intention to repurchase a product or to continue using a service, is one of the representative theories widely used in consumer behavior literature (Anderson and Sullivan, 1993; Oliver, 1980, 1993). The predictive power of ECT has been demonstrated in a wide range of products or services in the context of service marketing; more specifically, consumer satisfaction and post-purchase behavior (Anderson and Sullivan, 1993; Dabholkar et al., 2000; Oliver, 1980, 1993; Patterson et al. 1997; Tse and Wilton, 1988). The post-purchase behavior, representatively repurchase, has been explained including but not limited to camcorder (Spreng et al., 1996), photographic products (Dabholkar et al., 2000), automobile (Oliver, 1993), and public transit (Fu et al., 2018), recently expanding to IT products and services such as e-books (Valvi and West, 2013), internet banking (Rahi and Ghani, 2019), and wearable devices (Gupta et al., 2021).

While satisfaction is considered the key to building and maintaining a loyal base for long-term consumers, ECT hypothesizes that satisfaction with products/services, which is determined by the consumer’s initial expectations level with the product/service and its discrepancies with actual performance, determines consumer's intention to repurchase (Thong et al., 2006). The theory posits that expectations about the product/service are developed by consumers even before the purchase, and after the actual purchase, experience from using the product/service is gained by the consumer; say, the discrepancies may arise between expectations and actual delivery. In this way, consumers can confirm or disconfirm their expectations before purchase, evaluate perceived performance against previous frames of reference, and then form an opinion of performance. Consumer expectations are confirmed when the product/service meets or exceeds expectations (Churchill and Surprenant, 1982).

Figure 1 below illustrates the key constructs and relationships in the framework of ECT (Bhattacherjee, 2001). Consumers form repurchase intentions as the following process:

1. Initial expectations for a particular product or service is formed by consumers before purchasing.
2. Product or service is accepted or used. After an initial consumption, a perception of its performance is formed.
3. The product or service's perceived performance is evaluated against original expectations, and it is determined to what extent users' expectations are confirmed.
4. A level of satisfaction is formed according to their confirmation and expectation level on which that confirmation was based.
5. Satisfied consumers form repurchase intentions, and dissatisfied users churn from subsequent use.



Source: Bhattacherjee (2001)

Figure 1. Key constructs and relationships in ECT

The concept of satisfaction was extended by Oliver (1980) to the context of consumption as “the summary psychological state that occurs when emotions related to disconfirmed expectations are coupled with consumer’s prior feeling about consumption experiences.” However, the psychological state associated with the cognitive evaluation of the discrepancy between expectation and performance (confirmation) and the resulting psychological state were underscored. As illustrated in Figure 1, ECT assumes that confirmation is directly related to perceived performance, while being inversely proportional to expectations. Lower expectations and/or high-performance leads to greater level of confirmation, which then positively affects customer satisfaction and continuity (Bhattacherjee, 2001). The opposite vice versa causes disconfirmation and dissatisfaction, which lead to discontinuance intention.

ECT is also theorizes expectations as an additional determinant of satisfaction, as they provide a criterion or reference level by which consumers can make evaluation decisions about a particular product/service. Supporting this association is Helson's (1964) adaption level theory, that claim that people perceive deviations from the "adjusted level" or reference level of stimuli based on the situational context, characteristics of the stimulus, and the psychological characteristics of the person experiencing the stimulus. Higher baseline levels or expectations tend to lead to higher satisfaction, and lower expectations lead to lower subsequent satisfaction.

However, ECT has some limitations. First, even when post-purchase expectations may differ from pre-purchase expectations, it ignores possible changes in consumer expectations depending on the consumption experience. While pre-purchase expectations are generally based on the opinions or information provided by the media, post-purchase expectations are tempered by the consumers' direct experience (Fazio and Zanna 1981). Second, conceptualization of expectations differs across studies regarding ECT. Some define expectations as "anticipated performance", which is prior consumption beliefs on overall performance of a product/service (Westbrook and Reilly, 1983), when others, as beliefs about the level of attributes of a product or service (Oliver and Linda, 1981). Therefore, there exists the need to define expectations by incorporating the post-purchase expectations that are updated by consumer's direct experience.

2.1.2 Studies on Consumer's Continued Use of IT Service

Extending ECT paradigm, an Expectation Confirmation Model (ECM) of continuity in information technology (IT) was proposed by Bhattacherjee (2001) based on the correspondence between decisions of individuals' continued usage and repeat purchase of IT. The study argued that continuance decisions of IT users are similar to repurchase decisions of customers, which are influenced by the initial decision (acceptance or purchase) and the initial usage experience (of service or product), and can potentially lead to subsequent reversal of the initial decision. ECT's assertion that satisfaction with using IT, followed by perceived usefulness, is the strong predictor of users' intention to continue using the service/product was also supported. As such, the study contributed to the literature by conceptualizing the construct of confirmation and validating its effects on IT continuance intention through satisfaction. Researchers have also drawn upon recent findings of cognitive psychology literature by conceptualizing two dimensions, IT self-efficacy and facilitating conditions, to comprise perceived behavioral control, and have linked these dimensions to continuance behavior and intention (Bhattacherjee et al., 2008). It was found that users who lack confidence in particular IT use ability (i.e., low self-efficacy) are more likely to not continue using it even if they are reasonably satisfied with their previous usage experience, than users with high self-efficacy.

Similarly, several studies synthesize the ECM to explain and predict users' continuance intention of certain services from IT. Thong et al., (2006) adopts extended ECM that incorporates beliefs about perceived usefulness, enjoyment and ease of use after service use to explain the continued usage behavior. The results show that post-adoption service beliefs play an important role in the continued use of IT. In particular, the effect of perceived ease of use on continuance intention was strongest for mobile internet use. Other studies examine factors that determine customers' repurchase intentions in online shopping (Chiu et al., 2009; Lu and Su, 2009). Chiu et al. (2009) found significant predictors of customers' repurchase intentions in online shopping as trust, perceived ease of use, usefulness, and enjoyment. To add, Lu and Su (2009) demonstrated that enjoyment, usefulness, and compatibility are significant predictors, whereas anxiety was an emotional barrier against using innovative systems. Likewise, Lee (2010) found that satisfaction had the greatest influence on the users' intention to in continuously using the e-learning system.

Other recent studies explore ECM to identify factors that influence intention to continue using mobile applications (Hsiao et al., 2016; Tam et al., 2020; Alalwan, 2020). Hsiao et al. (2016) incorporated a customer value perspective to identify the factors that influence the intention to continue using social apps, which was driven by users' satisfaction, intimate connection with others, and hedonic motivation to use. Tam et al. (2020) demonstrated that the key drivers of persistent mobile app intentions using ECM along with the extended unified theory of acceptance and use of technology (UTAUT2). In terms of e-commerce platforms, Alalwan (2020) confirmed that hedonic motivation has a significant impact on both e-satisfaction and continued intention in mobile meal ordering apps. These results are due to the fact that functional benefits (i.e., ease of use and performance expectancy) in the consumer environment are sufficient in themselves to ensure customer satisfaction (Venkatesh et al., 2012). Therefore, it is consistently argued that in the decision of consumers to either adopt or reject new products and innovations are primarily shaped by psychological and hedonic benefits (Brown & Venkatesh, 2005; Davis et al., 1989; Van der Heijden, 2004).

The nature of these studies on consumers' continued use of services; however, are cross-sectional, therefore cannot accurately capture how customers' perception of product/service could change over time. Therefore, a longitudinal study is needed to discover how customers may adjust service expectations, perception, and satisfaction over time based on experience, and how they, in turn will affect consumer behaviors.

2.1.3 The GAP Model of Service Quality

The GAP service quality model first proposed by Parasuraman et al. (1985) to help organizations understand the factors that influence customer satisfaction. It is specifically used to understand various deviations, the "gaps", that occur in the process of providing services to potential customers by identifying the entire service delivery process and the gaps between processes. Suggesting that organization executives do not always know the characteristics that define high quality to consumers, characteristics a service must have to meet consumer needs, and levels of performance required to deliver high quality service, such lack of understanding affect consumer perception of quality.

According to the GAP Model, the five gaps that affect consumer's evaluation of service quality in the service provision process are as follows:

1. **Knowledge Gap**: The difference between expectations of consumers for service and the company's service provision. The gap may arise from lack of management's market research and therefore not knowing exactly what consumers want or need.
2. **Design (Policy) Gap**: The difference between perception of management and/or understanding and the actual specification of the consumer experience. When the management may correctly understand customer needs but fails to set performance standards. Administrators should ensure that the organization is defining the level of service that it deems necessary.
3. **Delivery Gap**: The difference between the service delivery policies and standards and the actual service delivery. It may be caused by a lack of workforce policies, employees who lack knowledge of the product, failure to match supply to demand, and lack of cohesive teamwork to deliver a product or service.
4. **Communication Gap**: The difference between what is promised to customers through advertisements and the actual delivery. This gap usually occurs when organizations exaggerate the deliverables to consumers, or share best practices to raise customer expectations and undermine customer perceptions.
5. **Customer Gap**: The difference between expectations and perceptions of consumers. Consumer expectations are shaped based on word-of-mouth, personal needs, and past experiences, but the actual perceived service quality may differ from expectations.

The concept of these gaps in service and their impact on customer satisfaction has long been researched. However, the GAP Model of service quality does not identify the gap between customer expectations and actual delivery in its impact on customer satisfaction.

As part of a decision strategy, consumers evaluate services based on expectations and relative experiences. As mentioned in the previous section, the ECT paradigm hypothesizes that consumer's pre-purchase expectation and the discrepancy between the expectation and the product/service's performance determine consumer’s satisfaction level with a product/service (Thong et al, 2006). In this regard, the concept of reference point was first introduced by Kahneman and Tversky (2013) in the prospect theory. Consumer behaviors are frequently observed as reference-dependent in economic situations, and therefore, it has been consistently emphasized in the behavioral economics and cognitive psychology literatures that relative levels of service may be more important than absolute levels in determining consumer preferences (Carson and Groves, 2007; DellaVigna, 2009; Hardie et al., 1993; Tversky and Kahneman, 1991). In particular, several studies report that the reference-dependent model perform better than standard choice modeling in exploring consumer behavior (Bateman et al., 2009; Kim et al., 2020). Therefore, to properly understand factors that influence consumer satisfaction of service, the gap between customer expectations, perceptions, and actual delivery, referred to as the “service gap” and the “perception gap”, need further elucidation in addition to the identified five gaps model of service quality.

2.1.4 Studies on Seasonality of Consumer Behaviors

Consumers' purchase of goods or use of services often show a specific pattern. Researchers have conducted various studies to study these patterns, some of which have tried to observe the consumer's natural seasonality in behaviors or to identify the consumer's usage behavior toward seasonal goods.

Seasonality is a phenomenon that affects several economic sectors, and among many, tourism a representative sector that is most identified by such seasonality (Fernández-Morales et al., 2016). While some studies suggest that the income level of tourists' countries of origin may have a direct impact on the seasonal behavior of visitors (Nadal et al., 2004), other studies suggest that while income changes inevitably affect aggregate demand, period of economic recovery may be associated with a decline in seasonal agglomeration (Duro and Turrión-Prats, 2019).

As such, previous studies have analyzed numerous factors involved in decision making and influencing seasonal trends. In analyzing the determinants, Hylleberg (1992) distinguishes three groups of factors: those related to the weather, specific events (religious events, festivals), and decisions tied to specific dates (school/company vacations, fiscal years). As seasonality analysis in tourism progressed, other explanatory factors were also presented. Reference has been made to the type of tourism products by the structure of tourism markets (Fernández-Morales et al., 2016), destinations and its potential for year-round availability and use (Cuccia and Rizzo, 2011; Martín et al., 2014), and seasonality and diversity of origin markets (Duro and Turrión-Prats, 2019). Martinez et al. (2020) specifically studied how the behavioral patterns of domestic and foreign tourists are influenced by the improvement of the economic conditions in terms of seasonal trends, in tourists arrivals.

Other groups of studies aimed to identify demand dynamics in the seasonal goods industry, where products are sold over limited seasons and have limited availability. In these markets, retailers often adopt dynamic pricing policies where initial sale prices are announced at the beginning of the season and price is reduced as the season progresses. In such market, strategic consumers find trade-off between buying early when prices are high but are available, and buying later in the season when the price is low but inventory risk is high (Soysal and Krishnamurthi, 2012). When dynamic pricing techniques are applied to the sales of seasonal product sales, sellers must consider the sales environment, including consumer behavior, product scarcity, and demand uncertainty (Aviv and Pazgal, 2008). Specifically, a study by Soysal and Krishnamurthi (2012) found that, in the context of fashion product, failure to account for consumer expectations of future availability or changes in seasonal total consumption utility can lead to biased demand estimates.

2.1.5 Studies on Online Platform Service Use

The online platform economy, recently characterized as the “future of work,” includes economic activities in which independent workers or vendors are provided with a platform to sell individual services or goods to consumers (Farrell and Greig, 2016). This includes labor platforms such as Uber that match consumers with providers who perform individual assignments or tasks, and capital platforms such as Airbnb and eBay that match consumers with providers who rent assets or sell goods. In digital economy, such platforms that bring together supply and demand are engines of growth, innovation, and competition, enabling businesses and consumers to take advantage of opportunities. In this regard, many of the existing studies have tried to understand consumer preferences for different online platform services and to develop marketing strategies.

Among studies conducted on the use of platform services, Dickinger and Mazanec (2008) focused on online hotel booking services, which is one of the most widely used online platforms. They investigated the hotel characteristics that influence consumer choice in an online booking environment by conducting an adaptive conjoint survey among 346 respondents. The results showed that the most important factors influencing online hotel booking were recommendations from friends and online reviews. More specifically, Chan et al. (2017) aimed to identify the impact of online reviews on consumers' buying decisions in hotel booking. The findings from experiments indicated that hotel booking intention were significantly affected by review valence (Chan et al., 2017). Similarly, Park et al. (2017) have found that popularity and consumer ratings had main effects on the booking intention. Others have investigated strategic consumer behavior in online hotel booking (Masiero et al. 2020; Alderighi et al., 2022). To derive consumers' preferences for free cancelations and nonrefundable rates in different scenarios, Masiero et al. (2020) conducted a discrete choice experiment. The result was that risk-seeking consumers were found to prefer the free cancelation rate, and that consumers' risk tolerance increases with the availability of automatic rebooking services. Based on data from Booking.com, Alderighi et al. (2022) added that dynamic pricing has a negative impact on consumers' perception of price fairness.

On other types of service, Suhartanto et al (2019) sought to identify different motives for using different types of online food delivery (OFD) services, with customer experience, restaurant search, ease of use, and listing being important antecedents for intention to use the service. In contrast, Suhartanto et al. (2019) found a direct effect of food quality on online loyalty. Similarly, Roh and Park (2019) showed that people's moral obligations in meal preparation can change attitudes that influence the decision to use OFDs. Chandrasekhar et al. (2019) conducted a comparative analysis of OFDs and consumer preferences and found that consumer perceptions play an important role in understanding their decision-making processes. To add, the study showed that consumers prefer uniqueness above all in terms of price, quality and delivery. Related to COVID -19, Habib et al. (2022) investigated the factors that favor consumers' online engagement (OCE) and platform preferences. The results showed that consumer self-concept and platform interactivity influenced OCE and platform preference during COVID -19. Meanwhile, Belarmino et al. (2021) compared antecedents of satisfaction before and during quarantine using ECT, and food quality, speed of service, ease of use, and belief confirmation during quarantine were found to be significant.

In the context of transport systems, Frei et al. (2017) conducted a representative analysis of demand for flexible, demand-responsive transit services in Chicago. They conducted a survey to determine consumer preferences and analyzed the data using choice models. The mode choice was set to three alternatives, such as (flexible) transit and car, while the attributes of these alternatives were walk time, wait time, and travel time, as well as cost, frequency, and number of transfers. The study found that bike-sharers or those who currently use active transportation or public transportation for commute were significantly more likely to choose flexible and traditional modes than car commuters.

Among the online platform services, e-hailing ride services (ERS), also known as ride-hailing and ride-sourcing, which are on-demand services that connect car owners and passengers via smartphones, are gaining popularity and changing the urban mobility landscape (Wang, 2019; Yan et al., 2020). Since stakeholder satisfaction with a service is a very important factor for the continuity of the service, a number of studies have recently been conducted to understand the objectives of various stakeholders of ERS and to find the equilibrium between them for the operation of sustainable transport systems. Existing studies on ERS can be broadly classified into followings: (1) understanding service adoption from the supply side and demand side perspectives and (2) identifying matching mechanisms and/or optimization of services. Specifically with respect to the ride-hailing platform, studies have mainly attempted to explain the service adoption from both the supply (drivers) and demand side (passenger) perspectives while trying to understand how ERS changes people's travel behavior.

As the popularity of ERS encourages the development of service that is not only reliable but is also consumer-oriented, numerous studies have been conducted to identify the factors that influence consumer intention and adoption of services (Arumugam et al., 2020; Nguyen-Phuoc et al., 2020). The studies have especially focused in developing countries where public transport systems are poorly developed due to lacking infrastructure, as ERS are increasingly seen as a substitute or complement to public transport. Representatively, Arumugam et al. (2020) studied consumer behavior toward ERS, which is rapidly becoming the preferred mode of public transportation in Malaysia. Referring to previous studies, they identified several factors that may influence user satisfaction and intention to use. As a result, subjective norms, perceived usefulness, perceived ease of use, compatibility, relative advantage, and safety were significant indicators of user satisfaction. Nguyen-Phuoc et al. (2020) focused on the factors affecting customer satisfaction and loyalty to ERS in Vietnam. They used partial least squares and structural equation modeling to identify the influencing factors from survey data of 559 ride-hailing passengers. The results show that perceived usefulness, sales promotion, and service quality directly influenced passenger satisfaction and loyalty.

Other studies analyzing the matching process in the sharing economy have mainly focused on demand-side heterogeneity. Ibrahim (2019) and Bai et al. (2019) considered consumer impatience with congestion as demand-side heterogeneity. Ibrahim (2019) also pointed out that the number of available workers is uncertain when the assumption that the firm controls whether and when workers work is no longer true. Bai et al. (2019) presented an analytical model that considers the number of participating agents and the demand rate. They found that the platform should charge a higher price when demand increases, but the price is not necessarily monotonic with the increase in provider capacity or waiting costs.

Studies that consider heterogeneity on the supply side develop optimal work schedules that incorporate flexible supplier self-schedules (Hu and Zhou, 2022; Chen et al., 2020).

While Wang et al. (2019) proposed a modified change-point model to derive adoption decisions and estimated changes in taxi drivers' driving behavior caused by mobile ride-hailing technology in Beijing, the optimal choice of trip selection was to favor longer trips than to aim for cruising time reduction to improve hourly earnings. Sun et al. (2020) proposed a new perspective on flexibility by considering the real-world choices of workers and drivers with respect to operating systems. Specifically, they modeled the matching process for taxi-hailing platforms and derived the optimal decisions for the platform and drivers using an approximate queuing analysis. Their analysis is based on real-world data from Didi Chuxing, which operates a number of matching systems: Inform and Assign. They found that the optimal radius is 1-3 kilometers and is lower during rush hour.

Recent studies in the field of management information and marketing have attempted to uncover the factors that bring about supply-side and demand-side satisfaction and that drive continued use in various platform environments. It is well defined in the marketing literature that ㅊontinued user satisfaction is considered key to building and retaining a loyal base of long-term customers. A consumer's intention to repurchase a product or continue to use a service depends primarily on satisfaction with the previous use of that product or service (Anderson and Sullivan, 1993; Oliver, 1980, 1993). There are some studies that identify the factors that influence consumer intention and adoption of services (Arumugam et al., 2020; Nguyen-Phuoc et al., 2020). However, the existing studies on the adoption behavior of platform services on the demand side do not adequately reflect the time-varying effects of continued use of the service.

2.2 Models with Time Effect

2.2.1 Fitting Data with Spline

Spline interpolation is an interpolation method where a special type of piecewise polynomial called a spline is used as an interpolant. Spline pieces together several functions in a principled way, that is, instead of fitting a single higher order polynomial to all values at once, spline interpolation fits a lower order polynomial to a small subset of values. It is often preferred to polynomial interpolation because the interpolation error can be small, even when using low-degree polynomials for splines. Spline interpolation also avoids the Lungi problem, which can cause oscillations between points when interpolating with higher order polynomials. Applications of these techniques include identifying trends in change for a given number. There also exists benefit of avoiding the problem of Runge's phenomenon, which can cause oscillations between points when interpolating with higher order polynomials. Application of these technique include identifying trends in change for a given number.

Suppose there are several points: $(x_0, y_0), (x_1, y_1), \dots (x_n, y_n)$. Then, we aim to find a smooth function $S(x)$ so that $S_k(x_k) = y_k$ for all k; more precisely, we seek k cubic polynomials $S_0, \dots, S_{k-1}$ for each x in the interval $[x_k, x_{k+1}]$. Such polynomials can be defined by Eq. (1).

$$S_k(x) = S_{k,0} + S_{k,1}(x - x_k) + S_{k,2}(x - x_k)^2 + S_{k,3}(x - x_k)^3 \cdots\cdots\cdots\cdots \text{Eq. (1)}$$

The main constraints for splines are as follows:

- **Constraint 1**: $S_k(x_k) = y_k$ (has to go through the data)
- **Constraint 2**: $S_k(x_{k+1}) = S_{k+1}(x_{k+1})$ (spline has to be continuous; two splines have to meet at a point)
- **Constraint 3**: $S'_k(x_{k+1}) = S'_{k+1}(x_{k+1})$ (1st derivative must also match)
- **Constraint 4**: $S''_k(x_{k+1}) = S''_{k+1}(x_{k+1})$ (2nd derivative (infliction point) must also match (extra smooth- strong constraint)

If there are $n + 1$ points (knots), then there exists n-intervals, and in each interval, there is a spline. The constraints can only be met with polynomials of degree 3 (cubic polynomial) or higher are used. The classical approach is to use the cubic splines, where there are n-intervals and $4 \cdot n$ unknowns. Then, there exists $4n - 2$ constraints: $n + 1$ constraint 1s, $n - 1$ constraint 2s as the splines have to meet at knots excluding those at the two ends, and likewise $n-1$ constraint 3s and 4s. Sometimes, additional constraints that the 2nd derivative at the two endpoints are zero are added (two additional constraints).

Because every interval is an individual cubic spline, there is no one global function, and there is a local fit. An important concept is that the constraint is used to estimate the polynomial within the interval. Intermediate values between knots can be identified through spline, but it is not good for extrapolation.

2.2.2 Varying Coefficient Models

Regression models are used to determine the numerical correlation between the predictors and response variables. However, when 'curvature' comes into play, the limitations of using linear regression models for estimation becomes apparent. There are two types of models used in the regression analysis: parametric and nonparametric. Parametric regression models are appropriate when the shape of the regression curve is known. Adopting a parametric regression curve necessitates access to other sources that provide point-by-point data. Otherwise, a nonparametric regression model can be employed as a constrained parametric regression will produce inconsistent results. In a nonparametric regression model, a curve fits into a function space, and the choice of function space is based on the characteristic of smoothness.

Likewise, smoothing splines with penalty (P-splines) can be used to estimate trends in expected values and time-varying regression coefficients modeled as curves or surfaces as part of a varying coefficient models (VCM) (Hastie and Tibshirani, 1993; Marx and Eilers, 1999; Ramsay and Silverman, 2003).

Hastie and Tibshirani (1993) first introduced VCM, where regression coefficients were allowed to interact with other variables, say, by varying as smooth functions of other variables to increase the flexibility of a linear regression model. The model suggests a linear relationship with the regressors, but the coefficients can change smoothly with the value of other variables (the effect modifiers). The simplest form is as follows:

$$E[y(t)] = \mu(t) = \beta(t)x(t) \cdots\cdots\cdots\cdots \text{Eq. (2)}$$

In Eq. (2), y and x are observed, β is estimated, and must change slowly with t. VCM assumes proportionality between y and x with a varying slope of the regression line. If B-spline basis B is assumed, then $\boldsymbol{\beta} = \boldsymbol{B\alpha}$ yielding $\boldsymbol{\mu} = \boldsymbol{XB\alpha}$ where $\boldsymbol{X} = diag(x)$. If a difference penalty is given on $\boldsymbol{\alpha}$, the equation has the structure of a P-spline with only the modified basis $\boldsymbol{XB}$. With a varying offset added, Eq. (2) can be expanded as Eq. (3) below in a form of an additive model, and if $\boldsymbol{\beta_0}$ is built with P-splines, it is called a P-GAM.

$$E[y(t)] = \mu(t) = \beta(t)x(t) + \beta_0(t) \cdots\cdots\cdots\cdots \text{Eq. (3)}$$

P-splines, which is smoothing splines with penalty, was originally introduced by Marx and Eilers (1999). A nonparametric technique previously used to flexibly estimate the effects of covariates, existing studies were mainly concerned with price response modeling in retailing (Steiner et al., 2007; Brezger and Steiner, 2008; Haupt et al., 2014). The idea of P-splines in the current context is to express the time-varying functions $f(\tau)$ in terms of a high-dimensional parametric basis and to add to the likelihood an appropriate penalty term corresponding to the regularization.

2.2.3 Discrete Choice Models with Time Effect in Attributes

A variety of choice models have been proposed to reflect the complexity of consumers' decision-making process, and a representative model is the dynamic discrete choice model. Here, consumers' choices are assumed to be the result of maximizing their present value of utility than the result of static utility maximization (Keane and Wolpin, 2009). Say, since consumers tend to prefer utility generated now or in the near future than utility generated in the far future, appropriate discounting should be applied to future utilities.

Its ultimate goal is to estimate the structural model of consumers' decision process, which includes intertemporal choices, based on Rust's (1987) framework. In the dynamic model, the agent's preference over period time t within discrete set $A = \{0,1,\dots,J\}$ and state vector s_{it} containing individual characteristics is represented as $\sum_{j=0}^{T} \beta^j U(\alpha_{i,t+j}, s_{i,t+j})$. Here, $U(a_{i,t}, s_{i,t})$ represents the current utility function, and β indicates a discounted factor. It is assumed that the evolution of future values of the state variables are affected by the decision at period time t. The agent is uncertain about the future values of the state, but can be represented as a Markov transition distribution function $F(s_{i,t+1}|a_{it}, s_{it})$. The vector of state variables s_{it} for each t is observed by the agent who chooses its action $a_{it} \in A$ that maximizes expected utility as follows:

$$E\left[\sum_{j=0}^{T-t} \beta^j U\left(a_{i,t+j}, s_{i,t+j}\right)|a_{it}, s_{it}\right] \cdots\cdots \text{Eq. (4)}$$

While the dynamic discrete choice model captures consumers' dynamic behaviors along with time preference, estimation requires computational burden.

Economists have also introduced several consumers' psychological consequences for consumers and integrated them into choice models, specifically for time preference. Largely, the literatures draw upon the modified choice experiments based on duration approach and time discounting. Duration approach allows for the estimation of the time trade-off among attributes by including duration in choice sets, and is commonly applied to estimate the value of a health state in health economics (Flynn, 2010; Norman et al., 2013; Bansback et al., 2014; Mulhern, 2017; Jonker, 2018). More specifically, a discount utility functions incorporating discounting function are used to reflect the discounting of value given to individuals over time as represented in following Eq. (5).

$$U_{njt} = (\boldsymbol{\beta}'_{\boldsymbol{n}} \boldsymbol{X}_{\boldsymbol{njt}})T_{ijt} + \epsilon_{ijt} \cdots\cdots \text{Eq. (5)}$$

As such, the discrete choice experiment under the duration approach is based on estimating consumer preferences through temporal tradeoffs by incorporating attributes of duration into the choice task. However, they have generally been analyzed as linear temporal preferences.

Other studies have included the “no-choice” or “past alternatives” options in discrete choice models to determine the timing of new project purchase. Haaijer et al. (2001) claim that a “no-choice” option should be included in conjoint choice experiments to enhance the model’s predictive power and to provide respondents with a more realistic choice situation. On the other hand, Dhar (1997) argued that respondents are more likely to choose the “no-choice” option when included in the choice set because it is more likely to stand out from the other alternatives.

2.3 Deep Learning Models for Data Prediction

2.3.1 Recurrent Neural Network (RNN)

A recurrent neural network (RNN) is a deep learning model for processing sequential data or time series. They are typically used in ordinal or temporal problems and are characterized by "memory" because they take in information from previous inputs to influence the current input/output. As shown in Figures 2 and 3, the hidden layer neural network H_t sees an input x_t and outputs y_t. It can be thought of as multiple copies of the same network, with each loop capable of passing information from one layer of the network to the next. It therefore connects previous information to the current task. This iterative module has a very simple structure like a single tanh layer of a neural network. The network is called the "recurrent" neural network because the repetition continues until the result value becomes optimized. The equations for RNN are shown in Eq. (6) to (9).

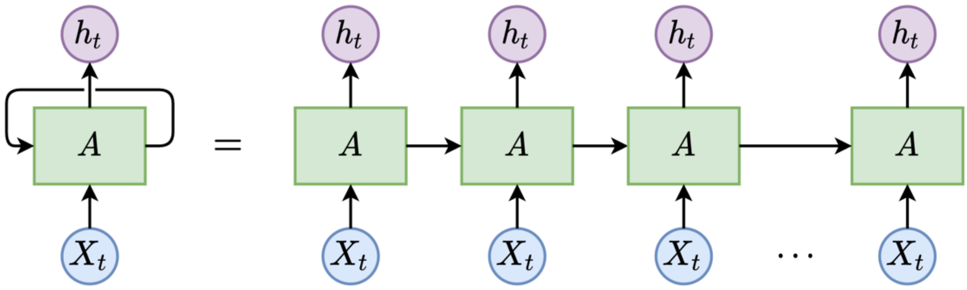


Figure 2. The unrolled recurrent neural network (RNN)

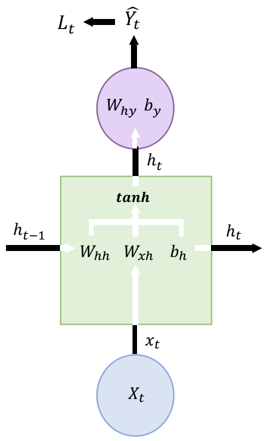


Figure 3. Internal structure of RNN

$$h_{t-1} = tanh(W_{hh}h_{t-2} + W_{xh}x_{t-1} + b_h) \cdots\cdots \text{Eq. (6)}$$

$$h_t = tanh(W_{hh}h_{t-1} + W_{xh}x_t + b_h) \cdots\cdots \text{Eq. (7)}$$

$$\hat{y}_t = W_{hy}h_t + b_y \cdots\cdots \text{Eq. (8)}$$

$$L_t = MSE = \sum(y_t - \hat{y}_t)^2/n \cdots\cdots \text{Eq. (9)}$$

In equations above, h_t is a hidden state at time t, W_{ij} is the weight from layer i to layer j, b_i is the bias in each layer, and L_t is the loss at time t. The model shares the weights and biases at all points in time and runs through the input data to output the results. The training of the RNN is iterated to minimize the loss due to the gradient descent of the loss function, taking into account the information about a particular previous time steps, while the weights are updated to find the optimal value. Such process, called the backpropagation through time (BPTT), can be expressed as follows, where η is the learning rate ranging [0, 1] (Chen, 2016):

$$UpdatedW_{xh} = ExistingW_{xh} - \eta \sum_{t=1}^{n} \sum_{k=0}^{n} \frac{\partial L_t}{\partial \hat{y}_t} \frac{\partial \hat{y}_t}{\partial h_t} \frac{\partial h_t}{\partial h_k} \frac{\partial h_k}{\partial W_{xh}} \cdots\cdots \text{Eq. (10)}$$

$$UpdatedW_{hh} = ExistingW_{hh} - \eta \sum_{t=1}^{n} \sum_{k=0}^{n} \frac{\partial L_t}{\partial \hat{y}_t} \frac{\partial \hat{y}_t}{\partial h_t} \frac{\partial h_t}{\partial h_k} \frac{\partial h_k}{\partial W_{hh}} \cdots\cdots \text{Eq. (11)}$$

$$UpdatedW_{hy} = ExistingW_{hy} - \eta \sum_{t=1}^{n} \frac{\partial L_t}{\partial W_{hy}} \cdots\cdots \text{Eq. (12)}$$

2.3.2 Long Short-Term Memory (LSTM)

LSTM refers to the structure of an artificial neural network used in artificial intelligence and deep learning fields. It is designed to enable long/short-term memory by compensating for the disadvantage that the existing RNN cannot contain information located far from the output even when it's needed, so called the "vanishing gradient" problem, as the gradient, the rate of weights update, disappears as the derivative value of tanh function with respect to h_t that is less than one is multiplied repetitively. LSTM is capable of learning long-term dependencies with chain-like structure (Hochreiter and Schmidhuber, 1997). Four interacting layers make up the repeating module in an LSTM, rather than a single neural network layer, as shown in Figure 4. In LSTM cells, the states are largely divided into two vectors: the h_t, short-term state, and the C_t, the long-term state. The process of each LSTM layer is summarized in Table 1.

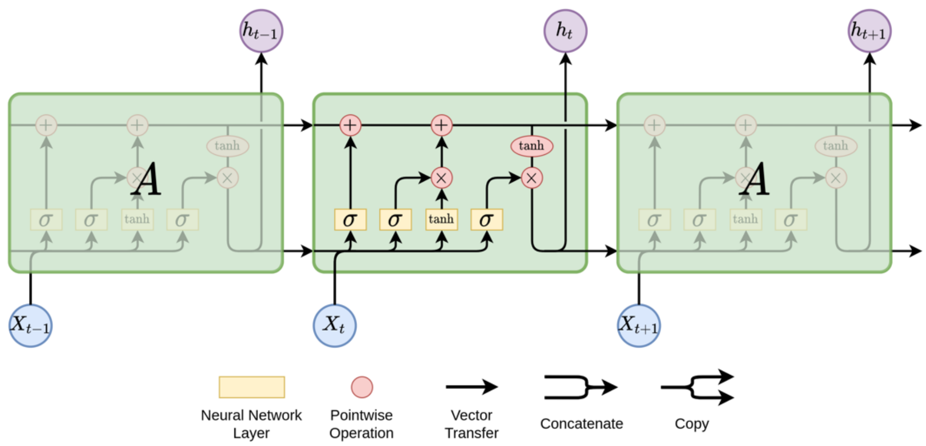
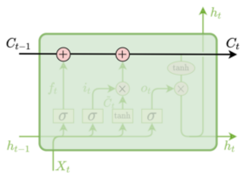
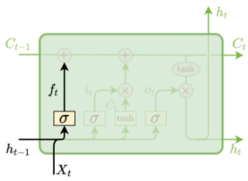
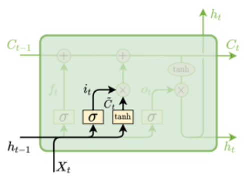
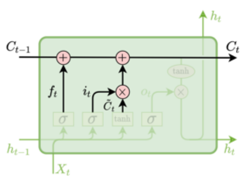
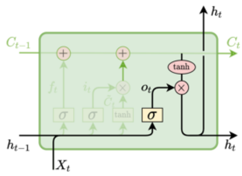


Figure 4. The repeating module in LSTM

In LSTM, forgetting and memory f_t, input i_t, inner cell state candidate $\widetilde{C}_t$, conveyance and inner cell state at time t, and output o_t are added to the RNN model. C_t contributes significantly to resolving long-term dependencies by penetrating all time points.

Table 1. Steps in LSTM

	Cell State - Responsible for allowing information to flow without change.
	Forget Gate - Decides what information to discard from the cell state through the sigmoid layer, also called the forget gate layer. $f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f)$
	Input Gate - Decides which of the new incoming information to be stored in the cell state. First, it goes through the sigmoid layer to decide which value to update. Next, a new candidate vector $\widetilde{C}_t$ is created in the tanh layer that could be added to the state. $i_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_i)$ $\widetilde{C}_t = tanh(W_C \cdot [h_{t-1}, x_t] + b_C)$
	Cell Status Update - After determining the information to be discarded and the information to be updated in the previous gate, C_{t-1} is updated into the new cell state C_t. $C_t = f_t \odot C_{t-1} + i_t \odot \widetilde{C}_t$

	Output Gate - Decides what information to send to the output based on the cell state, but filtered. - First, the output is determined by putting the input data in the sigmoid layer, then put the cell state in the tanh layer, multiply it with the output of the sigmoid layer, and export it as an output. $o_t = \sigma(W_o[h_{t-1}, x_t] + b_o)$ $h_t = o_t \odot \tanh(C_t)$

2.3.3 Applications of Deep Learning in Consumer Studies

While many studies have used quantitative approaches to capture diverse consumer preferences, with the expansion of deep learning applications, recent studies have attempted to use them in explaining consumer behaviors. In terms of product design, the study of Burnap and Hauser (2018) developed a deep learning approach to predict the "design gap", representing product designs that are not yet in markets but are highly preferred and are feasible to be built under existing constraints. By comparing design gaps retroactively predicted based on U.S. automotive market data with known successful designs in practice, the authors suggested that such an approach could identify market opportunities relatively early. Similarly, Bhat et al. (2019) used a machine learning-based 3D model quality assessment algorithm that mimics user preferences to select high-quality 3D models from a set of generated models. The results show that human intervention is not required in subjective quality analysis after learning user preferences, as the algorithm mimics user selection with 90% accuracy.

In the field of marketing, study of Yu (2021) advanced the concept and future development of interactive marketing by leveraging the power of natural language processing (NLP) to identify the interactions between marketers and consumers. Based on content generated by consumers and by marketers in the tourism industry, this study revealed the interaction relationships between them. In describing consumer behavior, RNNs were used help predict user changes (i.e., churn) using recorded user behavior in online card games (Xi et al., 2019). Khotimah and Sarno (2019) utilized LSTM using word embeddings from online reviews in a hotel service for sentiment classification. Interactive LSTM has also been used to predict sentiment from reviews of online restaurant (Luo and Xu, 2021). Furthermore, Oh et al. (2022) explored hospitality industry's consumer satisfaction by combining deep learning techniques with expectation-confirmation theory. The model achieved 83.54% accuracy in predicting customer satisfaction with hotel service using comments on hotel reviews, hotel information, and images. However, no studies have yet incorporated the deep learning techniques to identify expectations and/or perceptions of consumers that are updated with repeated use of services.

2.4 Limitations of Previous Literature and Research Motivation

Studies have attempted to identify factors that affect consumer satisfaction and behaviors in the context of IT services. However, existing studies utilize static utility models that inadequately reflect the time-varying effects of continued use of the service to explain consumer behaviors. Therefore, a dynamic longitudinal study is needed to uncover how consumers could adapt their experience and perception, and how they, in turn will affect consumer satisfaction and/or behaviors.

This dissertation therefore aims to fill this void by implementing a dynamic utility model to explain consumer behavior in the platform economy, where there is repeated use of service. As an empirical study, we aim to identify time-varying effects of covariates in explaining consumers' use of platform service by first incorporating accumulated experience effects rising from repeated use of service. In the second part of the study, we observe temporal effects in consumers' usage behaviors through semiparametric modeling. The study is a meaningful addition to the literature that aims to understand platform service adoption from the demand-side perspective and provide important implications that can be used in practice when considering strategies to retain and improve the loyalty of ride-hailing platform users.

Chapter 3. Methodology

3.1 Methodological Framework

In this dissertation, both linear regression model (generic model) and function with time-varying parameters are used for analysis. More specifically, a flexible regression model, penalized splines which is a flexible but parsimonious nonparametric smoothing method, is used for data-driven estimation time-varying effects on repeated service use behavior. In the model, a unified approach is used to determine the flexible functions as well as the corresponding degrees of smoothness. It can mimic state-space approaches with random-walk parameter dynamics and allows for time-varying effects of covariates.

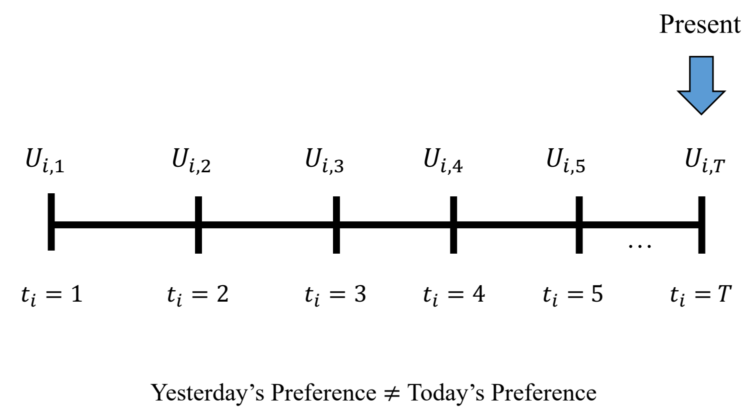


Figure 5. Time-varying preference of users

The semiparametric models are employed under the assumptions that people's preference on service may vary over time. The preferences of service users are captured by the parameter coefficients, which is assumed to be constant (and parametric) in the linear additive utility function, and non-constant (and nonparametric) in the semiparametric approach. Then, we construct the model under the assumption that people's preference on service is influenced by past usage experiences and that users make experience-based decisions, which emanates from direct or vicarious reinforcements that were received in the past. The conceptualizations of these assumptions are illustrated in Figures 5 and 6.

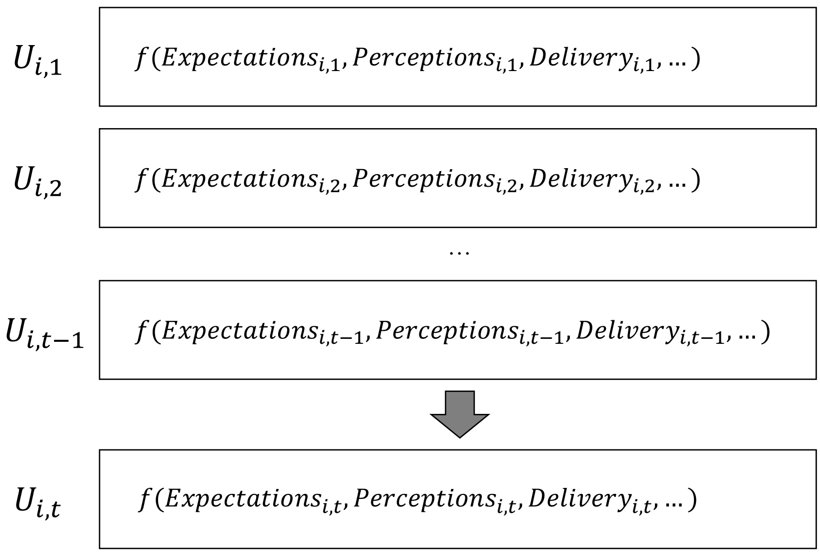


Figure 6. Accumulated experience and utility formation

3.2 Model Specification

3.2.1 Generic Model

The linear additive (indirect) utility that consumer n obtains from service J can be represented as follows:

$$U_{njt} = V_{njt} + \epsilon_{njt} = \alpha_j + \sum_{k=1}^{K} \beta_k \cdot x_{knjt} + \epsilon_{njt} \quad \text{Eq. (13)}$$

The utility in Eq. (13) is composed of the representative utility, V_{njt}, and the stochastic term, ϵ_{njt}. In this model, V_{njt} is defined as a linear function of K observable attributes x_{knjt} with corresponding coefficients β_k that represent effects of K independent variables. α_j represent intrinsic utilities of service J, also referred to as average effect of all unobserved variables in the model, and the random error term ϵ_{ijt} captures unobserved influences not covered by the data.

3.2.2 Functions with Time-Varying Parameters

The generic model (linear regression model) completely ignores possible temporal dependencies of individual usage behavior across successive usage occasions. Therefore, coefficients are assumed to be constant over the entire observation period, and each individual's observations over time are treated as independent. Time dependence can be introduced for both the intercept and the covariate effect to account for the time-varying parameters of the linear additive utility function.

First, time-varying intercepts is described by replacing the intercept parameter α_j with (a priori unknown) smooth time-dependent functions $f_j^{\alpha}(\tau)$ leading to the following regression function:

$$U_{njt} = f_j^{\alpha}(\tau_{nt}) + \sum_{k=1}^{K} \beta_k \cdot x_{knjt} + \epsilon_{njt} \cdots\cdots \text{Eq. (14)}$$

Then, the intercepts can account for changes in intrinsic utility over time (τ), which can be caused by long-term or short-term fluctuations. It is worth noting that the service use occasion t for each individual n occurs at a particular time τ_{nt}. In the empirical application of this dissertation, τ indexes days, and in most cases, t^{th} usage occurs at different τ for different individuals. Nevertheless, an individual may use the service multiple times within the same τ, and multiple users may use the service within the same τ, where the value of $f_j^{\alpha}(\tau)$ is identical.

Further, the covariates' time-varying effects can likewise be allowed by incorporating smooth time-dependent functions $f_j^{\beta}(\tau)$ to replace K time-constant effects β_k, leading to the following utility function:

$$U_{njt} = f_j^{\alpha}(\tau_{nt}) + \sum_{k=1}^{K} f_k^{\beta}(\tau_{nt}) \cdot x_{knjt} + \epsilon_{njt} \cdots\cdots \text{Eq. (15)}$$

Accordingly, it is possible to investigate whether the effects of marketing measures (such as price) or behavioral covariates that vary with time (such as reference price) change over time. Since the unknown time-varying function is modeled nonparametrically with a penalized spline (P-spline) for both the intercept and covariate effects, and the error term follows a parametric distribution as before, the utility functions from Eq. (14) and Eq. (15) are called semiparametric models.

3.2.3 Smoothing Splines and Penalized Regression

Nonparametric regression models are used to estimate a regression curve that depends only on the observed data. Functional estimation could be done using higher order global polynomials or by using splines which can be specified by choosing a set of knots. A spline, also called a regression spline, is a segmented polynomial model whose segment properties provide more flexibility than typical polynomial models. In a more locally oriented approach, the use of splines focuses on data patterns that have different specifications in one sector than in another. The spline regression model is adequately fitted to the local specification of the data (Hardle, 1990).

Smoothing splines with penalty (P-splines), first introduced by Marx and Eilers (1999), has previously been used as a nonparametric technique mostly concerned with price response modeling to flexibly estimate effects of retail covariates (Steiner et al., 2007; Brezger and Steiner, 2008; Haupt et al., 2014). The idea of P-splines in the present context is to represent the time-varying functions $f(\tau)$ in terms of a high-dimensional parametric basis and to add to the likelihood an appropriate penalty term for the same of regularization.

Assuming that for some continuous bivariate distribution $F_{X,Y}$ there exists an independent sample of n observations $(x_i, y_i) \sim F_{X,Y}$, we consider the following nonparametric regression model:

$$y_i = f(x_i) + \epsilon_i \quad \cdots\cdots \text{Eq. (16)}$$

, where $f(\cdot)$ is some unknown *smooth* function, and $\epsilon_i \sim iid\ (0, \sigma^2)$ are independently and identically distributed error terms with mean zero and variance σ^2, implying that $f(x_i)$ is the conditional mean of y_i given x_i. The objective is to estimate the unknown function $f(\cdot)$ from the data sample.

When we fit a curve to a data, the goal is to find a function that minimizes the residual sum of squares (RSS), which is defined as $\sum_{i=1}^{n}\left(y_i - f(x_i)\right)^2$. However, if there exist no constraints for $f(x_i)$, RSS can simply be made to equal 0 by selecting a curve that interpolates all of the y_i. However, such curve is overly flexible and can cause problems with overfitting.

Therefore, a smoothing spline that minimizes the penalized least squares function f_λ, which is represented in Eq. (17), is used to estimate $f(\cdot)$.

$$f_\lambda = \min_{f \in \mathcal{H}} \frac{1}{n} \sum_{i=1}^{n}\left(y_i - f(x_i)\right)^2 + \lambda J_m(f) \quad \cdots\cdots \text{Eq. (17)}$$

Here, $(\sum_{i=1}^{n}\left(y_i - f(x_i)\right)^2$ is a loss function, otherwise called the residual squares, that allows $f(x_i)$ to fit well to the data (measuring closeness to the data), and $\lambda J_m(f)$ is a penalty term that penalizes the curvature of $f(x_i)$. $\lambda > 0$ is the smoothing parameter (roughness coefficient) that controls the influence of the penalty. In other words, it controls the trade-off between too much flexibility (λ being too small) and sufficient smoothness (λ being too big). $J_m(f) = \int |f^m(z)|^2 dz$ quantifies the lack of parsimony of the functional estimate. $f^m(\cdot)$ denotes the m th derivative of $f(\cdot)$, say, with $f^2(z)$ representing the change of slope at point z. Therefore, $f^m(z)$ is a measure of roughness of the function. It is particularly notable that in most cases of smoothing splines, $J_m(f) = \int |f''(z)|^2 dz$ is used as a roughness penalty. Lastly, $\mathcal{H} = \{f: J_m(f) < \infty\}$ is the space of functions with square integrable m^{th} derivative.

3.2.3.1 Influence of the Smoothing Parameter

As mentioned previously, the goal is to find a function that minimizes the penalized RSS, which is defined in Eq. (5). The first and the second term measures the closeness to the data and penalizes the curvature of the function, respectively. Then, the two extreme cases are as follows:

- If $\lambda = 0$, $f(\cdot)$ is any function that interpolates the data.
- If $\lambda = \infty$, $f(\cdot)$ is the least squares fit for regression that satisfies $J_m(f_\lambda) \approx 0$.

That is, the closer λ is to 0, the less influence penalty has on the least squares function. Thus, for very small values, the function estimate f_λ essentially minimizes the residual sum of squares. On the other hand, the penalty has a greater effect on the penalized least squares function as λ approaches ∞. This means that the function estimate f_λ for very large values of λ must essentially have a penalty of zero that satisfies $J_m(f_\lambda) \approx 0$. As λ increases from 0 to ∞, the function estimate f_λ for the penalty function $J_m(\cdot)$ becomes smoother. The goal is to find the λ that gives the "right" degree of smoothing for the function estimate. An example of smoothing splines with varying smoothing parameters is represented in Figure 7.

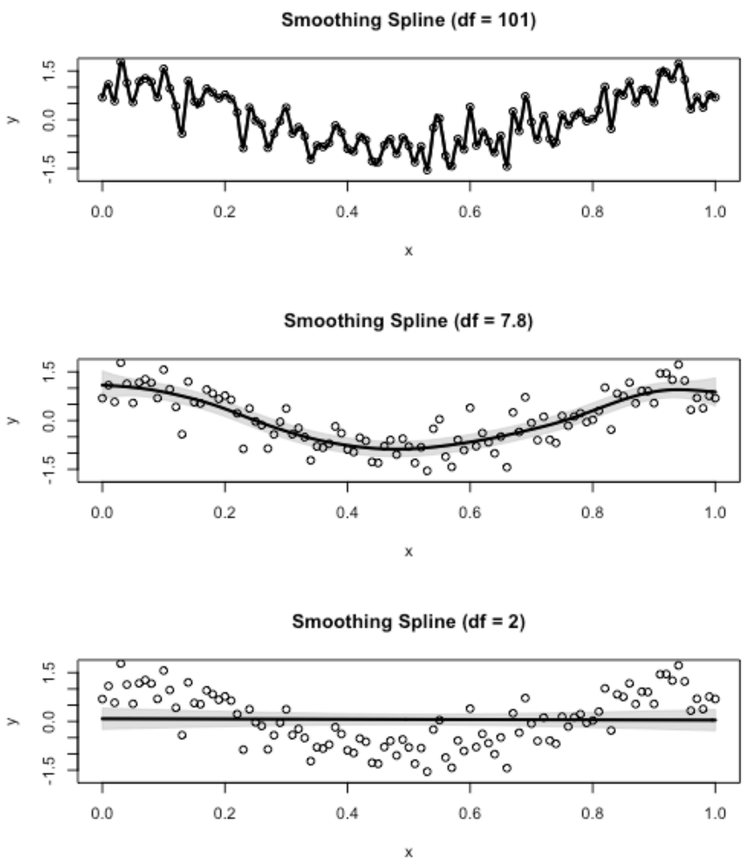


Figure 7. Influence of changing smoothing parameter (λ)

3.2.3.2 Influence of Penalty Order

In the empirical context of the dissertation, it is assumed that the unknown function $f(\tau)$ can be approximated by B-spline basis functions that has equally spaced intervals within the given time horizon. This leads to:

$$f(\tau) = \sum_{m=1}^{M} \gamma_m \cdot B_m(\tau, \delta) \quad \cdots\cdots\cdots \text{Eq. (18)}$$

, with $B_m(\tau; \delta)$ representing m^{th} B-spline basis functions of degree δ. γ_m is the unknown regression coefficient that is to be estimated from least squares for the m^{th} B-spline basis function (De Boor, 2001). The use of cubic splines is a fairly standard initial point for smoothing in generalized additive models (GAMS); however, the resulting functions may turn out too smooth in some situations (Wood, 2017). Eilers et al. (2015) therefore suggested to ensure enough flexibility for the unknown functions by using a relatively large number of intervals, allowing for the short-term fluctuations in consumers' service usage behavior in the context of our empirical study.

While a suitable penalty term can be derived from squared r^{th} order derivative, the derivative penalty with a roughness penalty based on first or second-order differences of adjacent regression coefficients γ_m can be approximated according to the B-spline theory, leading to the penalty terms of $\lambda \sum_{m=2}^{M} (\gamma_m - \gamma_{m-1})^2$ or $\lambda \sum_{m=3}^{M} (\gamma_m - 2\gamma_{m-1} - \gamma_{m-2})^2$, respectively. For statistical inference, the difference penalties can be represented in terms of the following quadratic forms: $\lambda\boldsymbol{\gamma}'\boldsymbol{P}_{(r)}\boldsymbol{\gamma}$, where the vector $\boldsymbol{\gamma}$ contains the M regression coefficients γ_m, and $\boldsymbol{P}_{(r)} = \boldsymbol{D}'_{(r)}\boldsymbol{D}_{(r)}$ corresponds to the penalty matrix constructed from the first or the second-order difference matrix as follows:

$$D_{(1)} = \begin{pmatrix} -1 & 1 & & & \\ & -1 & 1 & & \\ & & \ddots & & \\ & & -1 & 1 \end{pmatrix} \text{ and } D_{(2)} = \begin{pmatrix} 1 & -2 & 1 & & \\ & 1 & -2 & 1 & \\ & & \ddots & \ddots & \\ & & 1 & -2 & 1 \end{pmatrix} \cdots\cdots\cdots \text{Eq. (19)}$$

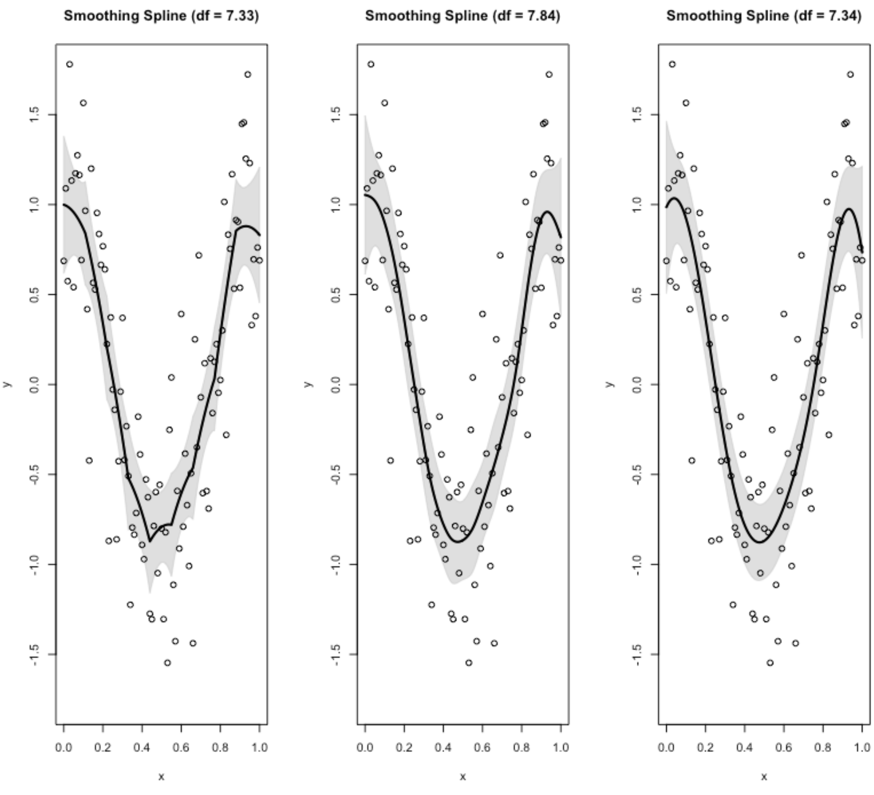


Figure 8. Influence of penalty order (δ) on smoothing splines

Likewise, B-splines with $\delta = 3$ is a cubic spline that penalizes the square of the second derivative of the function (twice continuously differentiable). A smooth cubic spline estimate $f(\cdot)$ using a piecewise cubic function connected at points are known as "knots," with two continuous derivatives that ensures a smooth estimate of the function and its derivatives. $\delta = 2$ results in a linear smoothing spline (a piecewise linear function), and $\delta = 4$ yields a quintic smoothing spline (a piecewise quintic function). An example of linear, cubic, and quintic smoothing splines is represented in Figure 8.

3.2.4 Estimation Method

The minimizer of Eq. (5) is finite-dimensional, although the criterion to be minimized lies over a Sobolev function space (function space for which the integral $\int |f^m(z)|^2$ is defined), an infinite-dimensional space. The Kimeldorf-Wahba representation theorem shows that the function $f \in \mathcal{H}$ minimizing the penalized least squares function has the following form:

$$f_\lambda(x) = \sum_{v=0}^{m-1} \beta_v N_v(x) + \sum_{u=1}^{r} \gamma_u K_1(x, x_u^*) \quad \cdots\cdots \text{Eq. (20)}$$

, where $\{N_v\}_{v=0}^{m-1}$ are unknown functions spanning the null space $\mathcal{H}_0 = \{f : J_m(f) = 0\}$. $K_1(\cdot,\cdot)$ is the reproducing kernel function that is known for the contrast space $\mathcal{H}_1 = \mathcal{H} \ominus \mathcal{H}_0$, and $\{x_u^*\}_{u=1}^{r}$ are the selected spline knots. $\boldsymbol{\beta} = (\beta_0, \dots, \beta_{m-1})^\mathrm{T}$ and $\boldsymbol{\gamma} = (\gamma_0, \dots, \gamma_{m-1})^\mathrm{T}$ are the unknown coefficient vectors of the basis functions. In an optimal solution, all n data points as knots are used. However, often, $r < n$ knots can be used to obtain good solutions, and in such cases, knots are typically placed at the quantiles of x_i.

If the Kimldorf-Wahba representor theorem is applied to the penalized least squares function, it can be rewritten as follows:

$$\frac{1}{n}\|\boldsymbol{y}-\boldsymbol{X\beta}-\boldsymbol{Z\gamma}\|^2+\lambda\boldsymbol{\gamma}^{\mathrm{T}}\boldsymbol{Q\gamma} \quad \cdots\cdots \text{Eq. (21)}$$

, where $\boldsymbol{y}=(y_1,\dots,y_n)^{\mathrm{T}}$ in the response vector, $\boldsymbol{X}=[N_v(x_i)]$ is a null space basis function matrix, $\boldsymbol{Z}=[K_1(x_i,x_u^*)]$ is the contrast space basis function matrix, and $\boldsymbol{Q}=[K_1(x_u^*,x_v^*)]$ is the penalty matrix, given Eq. (11), due to the reproducing property of the kernel function.

$$J_m(f_\lambda)=\sum_{u=1}^{r}\sum_{v=1}^{r}\gamma_u\gamma_v K_1(x_u^*,x_v^*) \quad \cdots\cdots \text{Eq. (22)}$$

The optimal basis function coefficients given λ can be written as Eq. (11), where $(\cdot)^{+}$ denotes the Moore-Penrose pseudoinverse. Because the coefficient estimates depend on the chosen smoothing parameter, meaning that different choice of λ result in different coefficient estimates, it is subscripted with λ.

$$\begin{bmatrix}\widehat{\boldsymbol{\beta}}_\lambda\\ \widehat{\boldsymbol{\gamma}}_\lambda\end{bmatrix}=\begin{bmatrix}\mathbf{X}^{\mathrm{T}}\mathbf{X} & \mathbf{X}^{\mathrm{T}}\mathbf{Z}\\ \mathbf{Z}^{\mathrm{T}}\mathbf{X} & \mathbf{Z}^{\mathrm{T}}\mathbf{Z}+n\lambda\mathbf{Q}\end{bmatrix}^{+}\begin{bmatrix}\mathbf{X}^{\mathrm{T}}\\ \mathbf{Z}^{\mathrm{T}}\end{bmatrix}\mathbf{y} \quad \cdots\cdots \text{Eq. (23)}$$

Then, the fitted values have the following form:

$$\hat{\mathbf{y}}_{\lambda} = \mathbf{X}\hat{\boldsymbol{\beta}}_{\lambda} + \mathbf{Z}\hat{\boldsymbol{\gamma}}_{\lambda} = \mathbf{S}_{y}\mathbf{y} \quad \text{Eq. (24)}$$

, where the smoothing matrix is as in Eq. (13). It is the smoothing spline analogue of the "hat matrix" $\mathbf{H} = \mathbf{X}\left(\mathbf{X}^{\mathbf{T}}\mathbf{X}\right)^{-\mathbf{1}}\mathbf{X}^{\mathbf{T}}$ in linear regression.

$$\mathbf{S}_{\lambda} = [\mathbf{X} \quad \mathbf{Z}] = \begin{bmatrix} \mathbf{X}^{\mathrm{T}}\mathbf{X} & \mathbf{X}^{\mathrm{T}}\mathbf{Z} \\ \mathbf{Z}^{\mathrm{T}}\mathbf{X} & \mathbf{Z}^{\mathrm{T}}\mathbf{Z} + n\lambda\mathbf{Q} \end{bmatrix}^{+} \begin{bmatrix} \mathbf{X}^{\mathrm{T}} \\ \mathbf{Z}^{\mathrm{T}} \end{bmatrix} \quad \text{Eq. (25)}$$

In a parametric regression model, the number of parameters (i.e. regression coefficients) are equivalent to the degrees of freedom (DF) of a model fit. Nonetheless, for smoothing splines, the number of coefficients could be equal or be greater than the number of observations n; therefore, such statement is not applicable.

In a nonparametric regression model, the effective (or equivalent) degrees of freedom (EDF) is defined as follows, where $\mathrm{tr}(\cdot)$ is a matrix trace function.

$$\nu_{\lambda} = \mathrm{tr}(\mathbf{S}_{\lambda}) \quad \text{Eq. (26)}$$

The EDF changes as a function of λ in a way that as λ approaches 0, the it approaches $m + r$ (null space dimension + number of knots), and as λ approaches ∞, it approaches m (null space dimension). As mentioned previously, the trace of the "hat matrix" in multiple linear regression is identical to the number of coefficients; therefore, the DF defined in the model above is a direct analogue of the DF defined in a multiple linear regression model.

3.2.5 Parameter Selection

3.2.5.1 Cross-Validations Method

The use of ordinary cross-validation (OCV), also referred to as leave-on-out cross-validation (LOO-CV), for model selection and assessment discovered by Allen (1974) and Stone (1974), and it was later suggested the use of OCV when fitting smoothing spline models (Wahba and Wold, 1975). The OCV can be used to find the λ that minimizes the following:

$$OCV(\lambda) = \frac{1}{n}\sum_{i=1}^{n}\left(y_i - f_\lambda^{[i]}(x_i)\right)^2 \cdots\cdots \text{Eq. (27)}$$

, where $f_\lambda^{[i]} \in \mathcal{H}$ is the function that minimizes the penalized least squares function that leaves out the i^{th} pair (x_i, y_i), as follows:

$$\eta_\lambda^{[i]} = \min_{f \in \mathcal{H}} \frac{1}{n}\sum_{j=1, j\neq i}^{n}\left(y_i - f(x_i)\right)^2 + \lambda J_m(f) \cdots\cdots \text{Eq. (28)}$$

In other words, Eq. (28) is the minimizer of the leave-one-out version of the penalized least squares function. Equation (27) by definition suggests that evaluating the OCV criterion for a given λ requires fitting the model n different times, once for each x_i. However, it can be shown that the OCV can be evaluated for a given λ using the results from the single model fit to the full sample of data, such as:

$$OCV(\lambda) = \frac{1}{n}\sum_{i=1}^{n}\left(y_i - f_\lambda(x_i) \Big/ 1 - s_{ii(\lambda)}\right)^2 \cdots\cdots \text{Eq. (29)}$$

, where $s_{ii(\lambda)}$ is the i^{th} diagonal element of $\mathbf{S}_\lambda$. Eq. (29), a computational form of the OCV criterion, is yielded by plugging in the following form of $\eta_\lambda^{[i]}$ into the OCV criterion in Eq. (27). Different weight are given to each observation for CV tuning, where the weights are defined as $w_i = \left(1 - s_{ii(\lambda)}\right)^2$, and the leverages satisfy $s_{ii(\lambda)} \in (0,1)$, that differs across observations.

$$\eta_\lambda^{[i]} = \eta_\lambda(x_i) - s_{ii(\lambda)}y_i \Big/ 1 - s_{ii(\lambda)} \cdots\cdots \text{Eq. (30)}$$

An improved version of OCV is the generalized cross-validation criterion (GCV) that equalizes the influence of observations on the smoothing parameter selection by replacing the leverages with their average value, $\frac{1}{n}\sum_{i=1}^{n} s_{ii(\lambda)} = v_\lambda/n$, as first suggested by Craven and Wahba (1978). In other words, GCV seeks to find λ that minimizes the following:

$$GCV(\lambda) = \frac{1}{n}\sum_{i=1}^{n}\left(y_i - f_\lambda(x_i)\right)^2 \Big/ \left(1 - \frac{v_\lambda}{n}\right)^2 \cdots\cdots \text{Eq. (31)}$$

, where $v_\lambda = \mathrm{tr}(\mathbf{S}_\lambda)$ is the EDF of the estimator η_λ. The GCV criterion is preferred over the OCV as it has desirable asymptotic properties assuming $\epsilon_i \sim iid\ (0, \sigma^2)$, especially when there exists replicate x_i scores in the sample (Li, 1987).

3.2.5.1 Information Criteria

Meanwhile, the usage of information criteria requires more assumption than the cross-validation based methods. If the error terms are assumed to be i.i.d. Gaussian, say, $\epsilon_i \sim iid\ N(0, \sigma^2)$ which implies that $y_i \sim ind\ N(f(x_i), \sigma^2)$, information criteria can be used to select the smoothing parameter λ.

The log-likelihood function given a sample of n independent observations has the following form:

$$l(\lambda, \sigma^2) = -\frac{1}{2\sigma^2}\left\{\sum_{i=1}^{n}\left(y_i - f_\lambda(x_i)\right)^2 - \frac{n}{2}\log(\sigma^2) - \frac{n}{2}\log(2\pi)\right\} \cdots\cdots\cdots \text{Eq. (32)}$$

, which depends on the smoothing parameter λ and the error variance σ^2. The error variance is known in most cases, therefore, the maximum likelihood estimate $\sigma_\lambda^2 = \frac{1}{n}\sum_{i=1}^{n}(y_i - f_\lambda(x_i))^2$ can be used. Then, by substituting σ_λ^2 for σ^2, the log-likelihood as a function of λ is yielded as in Eq. (21). It can be noted that the following equation only depends on λ through $\log(\sigma_\lambda^2)$ with other terms constant.

$$\tilde{l}(\lambda) = l(\lambda, \sigma_\lambda^2) = -\frac{n}{2} - \frac{n}{2}\log(\sigma_\lambda^2) - \frac{n}{2}\log(2\pi) \cdots\cdots \text{Eq. (33)}$$

To select smoothing parameters, the Akaike Information Criterion (AIC) and the Bayesian Information Criterion (BIC) proposed by Akaike (1974) and Schwarz (1978) can be used by adding a penalty to the log likelihood $\tilde{l}(\lambda)$. AIC aims to select a model whose BIC is similar to AIC but uses different weights in the penalty to select the model that loses the least information about the unknown true data generation process. The two criteria seek to find λ that minimizes the following:

$$AIC(\lambda) = -2\tilde{l}(\lambda) + 2v_\lambda \cdots\cdots \text{Eq. (34)}$$

$$BIC(\lambda) = -2\tilde{l}(\lambda) + \log(n)v_\lambda \cdots\cdots \text{Eq. (35)}$$

3.2.5.2 Maximum Likelihood

The maximum likelihood (ML) approach exploits the computational relationship between a penalized spline and a linear mixed effects model (Wahba, 1985; Wang, 1998; Ruppert et al., 2003). The underlying approach is similar to the Bayesian confidence intervals (CI) except that the null space coefficients are treated as fixed effects.

Let us assume that $\boldsymbol{\gamma} \sim N\left(\mathbf{0}, \frac{\sigma^2}{n\lambda}\mathbf{Q}^{-1}\right)$ and $\boldsymbol{\epsilon} \sim N(\mathbf{0}, \sigma^2\mathbf{I})$, where $\boldsymbol{\epsilon} = (\epsilon_1, \dots, \epsilon_n)^\mathrm{T}$ is the error vector and γ is independent of ϵ. Then, the response vector is $\mathbf{y} \sim N(\boldsymbol{X\beta}, \sigma^2\boldsymbol{\Sigma}_\lambda)$, where $\boldsymbol{\Sigma}_\lambda$ is part of the covariance matrix that is λ dependent. The null space representation and the contrast space representation contains the fixed and random effects, respectively.

$$\boldsymbol{\Sigma}_\lambda = {}^{1}/_{n\lambda}\left(\mathbf{Z}\mathbf{Q}^{-1}\mathbf{Z}^{\mathrm{T}} + \mathbf{I}\right) \cdots\cdots \text{Eq. (36)}$$

Then, the log-likelihood function has the following form with independent sample of n observations, where $\mathbf{r} = \mathbf{y} - \mathbf{X}\boldsymbol{\beta}$.

$$L(\lambda, \sigma^2) = -\frac{1}{2}\{\sigma^{-2} r^T \boldsymbol{\Sigma}_\lambda^{-1} r + log(|\boldsymbol{\Sigma}_\lambda|) + nlog(\sigma^2) + nlog(2\pi)\} \cdots\cdots \text{Eq. (37)}$$

The ML estimate for σ^2 has the following form since in most cases, σ^2 is unknown:

$$\sigma^2_{\lambda(ML)} = \frac{1}{n}\mathbf{r}^{\mathrm{T}}\boldsymbol{\Sigma}_\lambda^{-1}\mathbf{r} \cdots\cdots \text{Eq. (38)}$$

Substituting Eq. (25) for σ^2 of the log-likelihood function produces the following ML criterion that depends on λ through $\boldsymbol{\Sigma}_\lambda$:

$$ML(\lambda) = -\frac{1}{2}\left\{n + log(|\boldsymbol{\Sigma}_\lambda|) + nlog\left(\mathbf{r}^{\mathrm{T}}\boldsymbol{\Sigma}_\lambda^{-1}\mathbf{r}\right) + nlog(2\pi/n)\right\} \cdots\cdots \text{Eq. (39)}$$

The restricted maximum likelihood (REML) estimation accounts for DF reduction due to estimation of m null space coefficients (Patterson, 1971), yielding the following log-likelihood:

$$R(\lambda, \sigma^2) = L(\lambda, \sigma^2) - \frac{1}{2}\left\{log\left|\mathbf{X}^{\mathrm{T}}\mathbf{\Sigma}_{\lambda}^{-1}\mathbf{X}\right| - mlog(2\pi\sigma^2)\right\} \cdots\cdots\cdots\cdots\cdots\cdots\cdots \text{Eq. (40)}$$

, implying that the REML estimate for σ^2 has the following form:

$$\sigma^2_{\lambda(REML)} = \frac{1}{n-m}\mathbf{r}^{\mathrm{T}}\mathbf{\Sigma}_{\lambda}^{-1}\mathbf{r} \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \text{Eq. (41)}$$

Substituting Eq. (29) for σ^2 of the log-likelihood function produces the following REML criterion in Eq. (30), where $\tilde{n} = n - m$ indicates the DF corresponding to $\sigma^2_{\lambda(REML)}$.

$$REML(\lambda) = -\frac{1}{2}\left\{\tilde{n} + log(|\mathbf{\Sigma}_{\lambda}|) + \tilde{n}log\left(\mathbf{r}^{\mathrm{T}}\mathbf{\Sigma}_{\lambda}^{-1}\mathbf{r}\right) + \tilde{n}log\left({}^{2\pi}/_{\tilde{n}}\right) + log\left(\left|\mathbf{X}^{\mathrm{T}}\mathbf{\Sigma}_{\lambda}^{-1}\mathbf{X}\right|\right)\right\}$$

$$\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \text{Eq. (42)}$$

Chapter 4. Simulation Study

4.1 Validation of P-spine Implementation

In this section, the proposed smoothing spline model with penalty, the P-splines, is implemented and verified using the R software. In particular, this section aims to show that implementation of nonparametric models is more suitable for fitting data with variation (i.e. time variance caused by long-term or short-term fluctuations) than the parametric models (i.e. linear additive function) where parametric coefficients are assumed to be constant.

For validation, a set of data is randomly generated according to the model in Eq. (4) using the following equation.

$$f_1(x_i) = 2 + 4 \cdot \sin(2\pi x) \quad \text{Eq. (43)}$$

The randomly generated data created based on a noise of the function in Eq. (43) is represented in Figure 9 below. It can be observed that the simulated data fluctuates with the values of x. For the generated data, P-spline model was fit using the R program's ss() function of npreg package. $r = 10$ knots were placed evenly across the range of the x_i scores. Amongst the smoothing parameter selection methods, GCV method was used as it was the default choice in the implemented program.

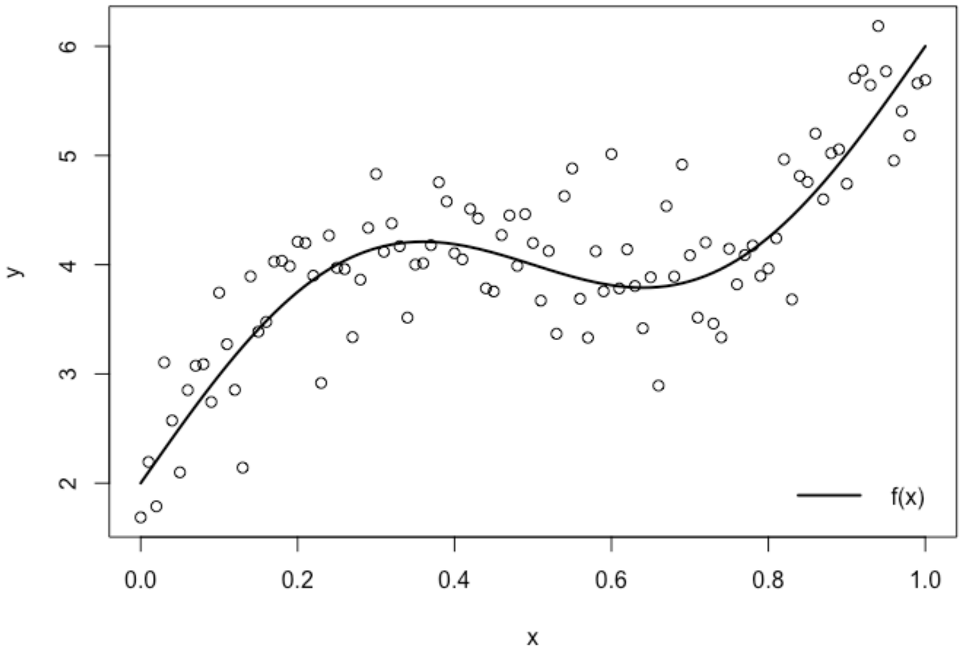


Figure 9. Randomly created data based on $f_1(x_i)$ for spline validation

Then, the smoothing spline with penalty (P-spline) was fit for function $f_1(x_i)$. The approximate significance of the parametric and nonparametric effects from P-spline fit for $f_1(x_i)$ is presented in Tables 2 and 3, respectively.

Table 2. Approx. significance of parametric effects from P-spline fit for $f_1(x_i)$

Parameter	Estimate	Std. Error	t-value	Pr(>$\|t\|$)
(Intercept)	4.052	0.0450	90.07	0***
x	3.688	0.2844	12.97	0***

*** p<0.001, ** p<0.01, * p<0.05, . p<0.1

Table 3. Approx. significance of nonparametric effects from P-spline fit for $f_1(x_i)$

Parameter	DF	Sums of Squares	Mean Squares	F-value	Pr(>F)
$f(x)$	4.432	17.60	3.9712	19.45	2.09e-12***
Residuals	94.568	19.31	0.2042	-	-

*** p<0.001, ** p<0.01, * p<0.05, . p<0.1

Note that the model has a coefficient of determination, R^2, of 0.7464 with DF of 94.57. Then, the coverage of the 95% Bayesian confidence interval (CI) for each smoothing parameter selection method can be calculated using the following Eq. (44), where $I\{\cdot\}$ denotes an indicator function.

$$Coverage = \frac{1}{n}\sum_{i=1}^{n} I\left\{a\left(\hat{f}_\lambda(x_i)\right) \leq f(x_i) \leq b\left(\hat{f}_\lambda(x_i)\right)\right\} \quad \cdots\cdots\cdots\cdots\cdots\cdots \text{Eq. (44)}$$

The upper bound and the lower bound for the 95% Bayesian CI is $a\left(\hat{f}_\lambda(x_i)\right) = \hat{f}_\lambda(x_i) - 1.96\hat{\sigma}_\lambda\sqrt{s_{ii(\lambda)}}$ and $b\left(\hat{f}_\lambda(x_i)\right) = \hat{f}_\lambda(x_i) + 1.96\hat{\sigma}_\lambda\sqrt{s_{ii(\lambda)}}$, respectively. Figure 10 shows the estimated functional relationship using the P-spline fit for $f_1(x_i)$, and the gray shaded area denotes the simulated coverage of the 95% Bayesian CI of the function. From the figure, it can be noted that the linear model fit does not completely fall within the Bayesian CIs, which suggests that the nonparametric model better fits the given data.

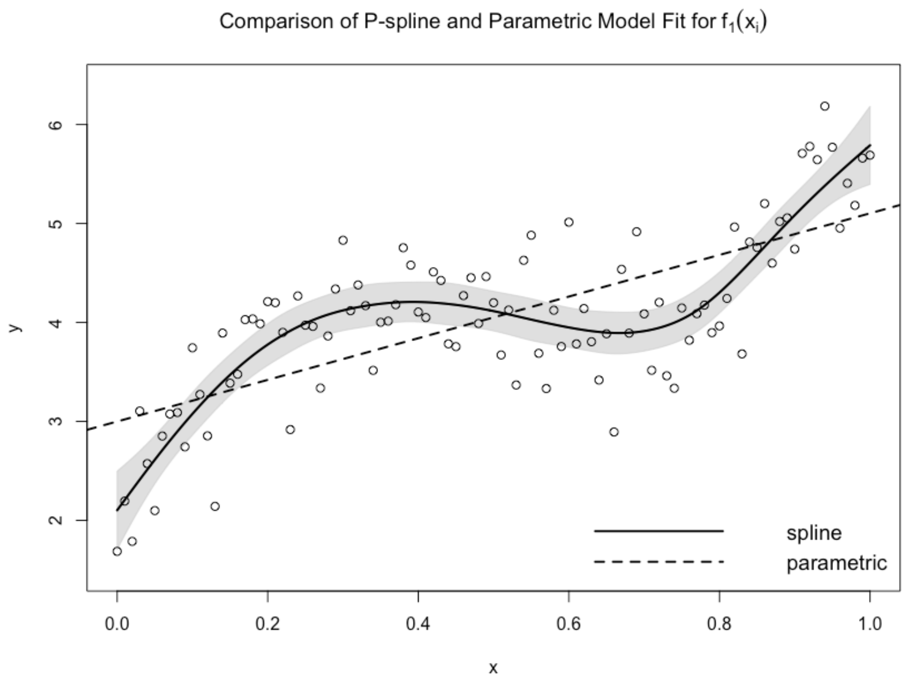


Figure 10. Comparison of P-spline and parametric model fit for $f_1(x_i)$

The root mean squared error (RMSE) of the estimator can be calculated using Eq. (45) to evaluate the performance of the estimation method. Smaller value of RMSE indicates better fit (or recovery) of the true mean function. By comparing the RMSE value of the spline and the parametric fit, it can be determined whether the nonparametric model is a better fit of the given data.

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(f(x_i) - \hat{f}_\lambda(x_i)\right)^2} \quad \text{Eq. (45)}$$

It is yielded that the RMSE value of the P-spline is 0.0836, whereas the RMSE value of the parametric model (linear fit) is 0.4537. It can also be noted that the RMSE value between P-spline and the parametric model fit is 0.4036. Because the value of P-spline is far smaller than that of the parametric model, it is once again validated that the nonparametric model should be preferred in fitting the given data that fluctuates with the values of x. The RMSE values for P-spline and parametric model fits for $f_1(x_i)$ are summarized in Table 4.

Table 4. RMSE value comparison for P-spline and parametric model fits for $f_1(x_i)$

	Between P-spline Fit and $f_1(x_i)$	Between Parametric Model Fit and $f_1(x_i)$	Between P-spline Fit and Parametric Model Fit
RMSE	0.0836	0.4537	0.4036

4.2 Functional Case Studies

For functional case studies, three types of data are generated for comparison of P-spline fit. First is a set of data that (1) strictly forms linear relationship, (2) does not fluctuate with time or fluctuates with small variation, and (3) fluctuates with time (big variation).

$$f_2(x_i) = 2x + 3 \quad \cdots\cdots \text{Eq. (46)}$$

$$f_3(x_i) = x^2 + x^3 + \sin(3x) \quad \cdots\cdots \text{Eq. (47)}$$

$$f_4(x_i) = \sin(2\pi x + 2) \quad \cdots\cdots \text{Eq. (48)}$$

The randomly generated data created based on a noise of the functions in Eq. (46), Eq. (47), and Eq. (48) are represented in Figures 11, 12, and 13, respectively. Again, for the generated data, P-spline model was fit using the R program via the ss() function. Figures 14, Figure 15, and Figure 16 shows the estimated functional relationship using the P-spline fits for $f_2(x_i)$, $f_3(x_i)$, and $f_4(x_i)$, and the gray shaded area denotes the simulated coverage of the 95% Bayesian CI of the function.

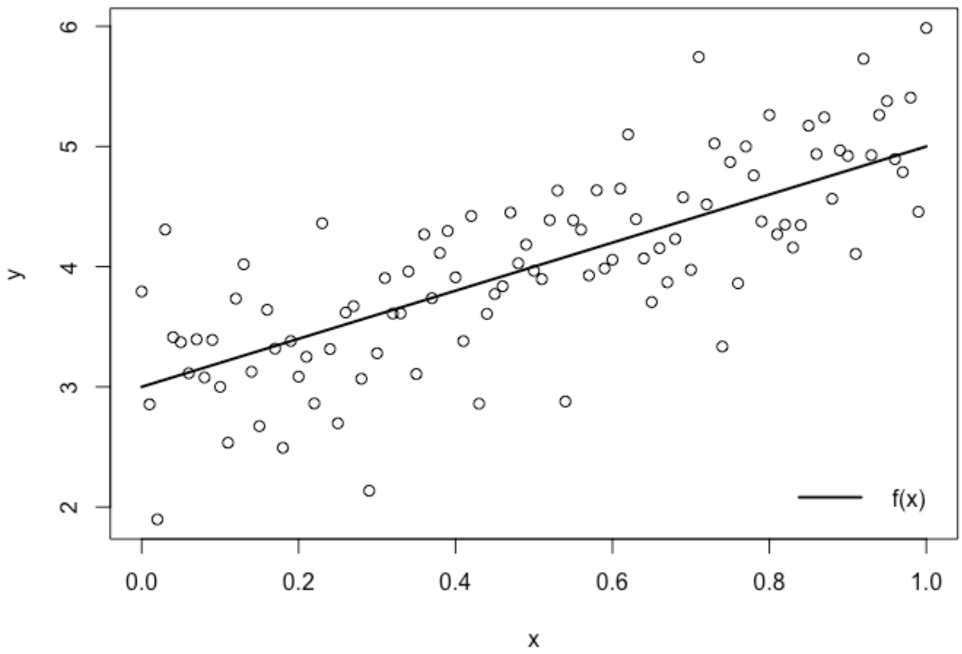


Figure 11. Randomly created data based on $f_2(x_i)$ for simulation

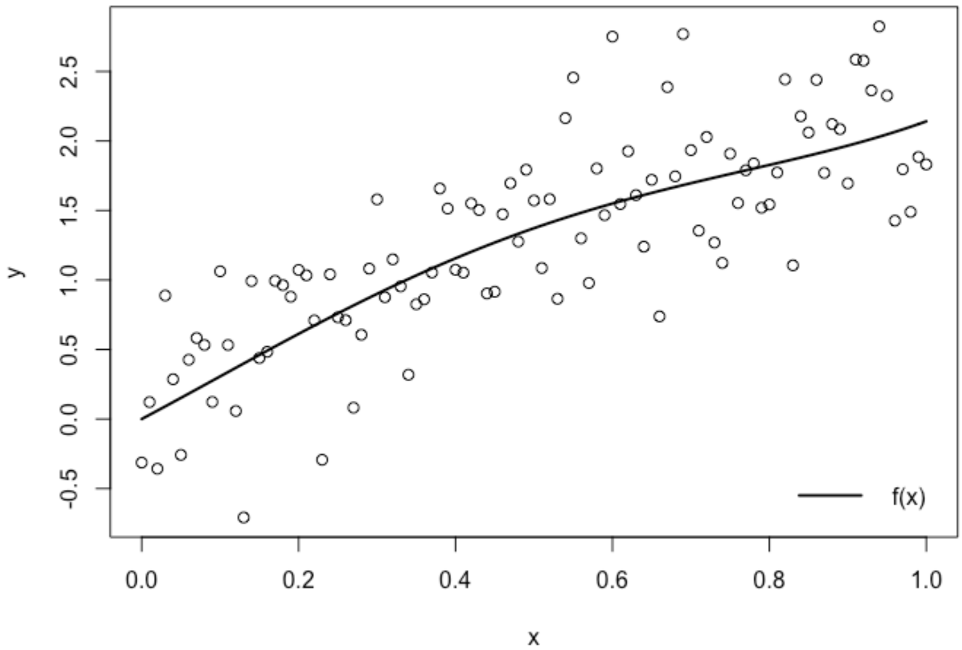


Figure 12. Randomly created data based on $f_3(x_i)$ for simulation

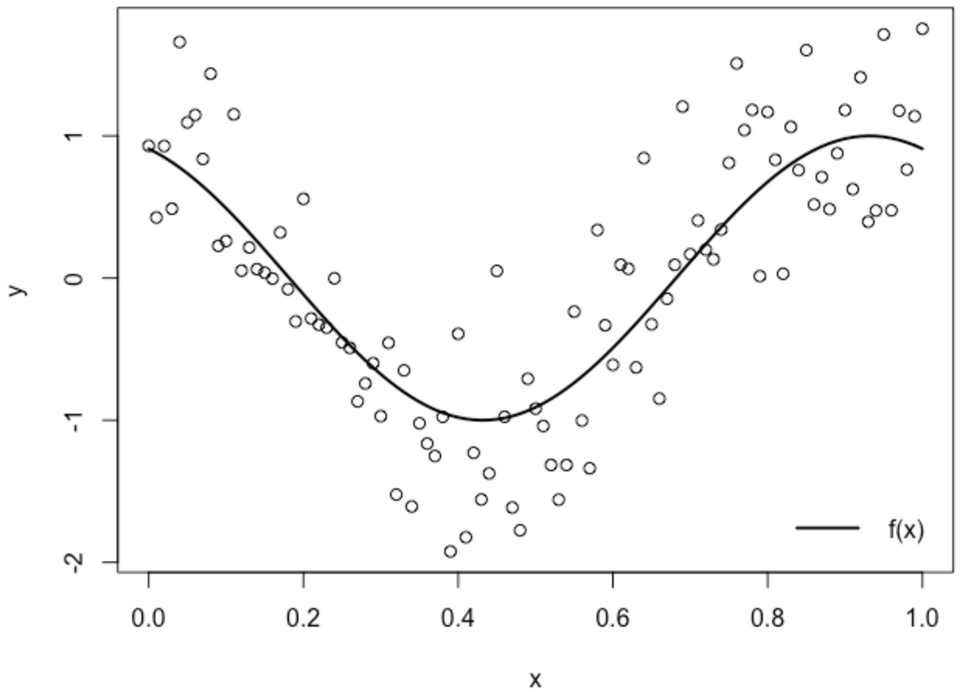


Figure 13. Randomly created data based on $f_4(x_i)$ for simulation

It is noticeable that in Figure 14 and 15, the parametric model fit falls completely within the Bayesian CIs, whereas in Figure 16, the parametric model fit does not, suggesting that the nonparametric model better fits the given data generated from function $f_4(x_i)$.

The root mean squared error (RMSE) of the estimator was calculated using Eq. (45) to evaluate the performance of the estimation method, and it was yielded that for $f_2(x_i)$, the RMSE value of the P-spline is 0.0453, whereas the RMSE value of the parametric model (linear fit) is 0.0451. The RMSE value between the P-spline and the parametric model fit is 0.0038. The results indicate that while the two fits are relatively similarly close to the true value, parametric model is a better recovery of the true mean function.

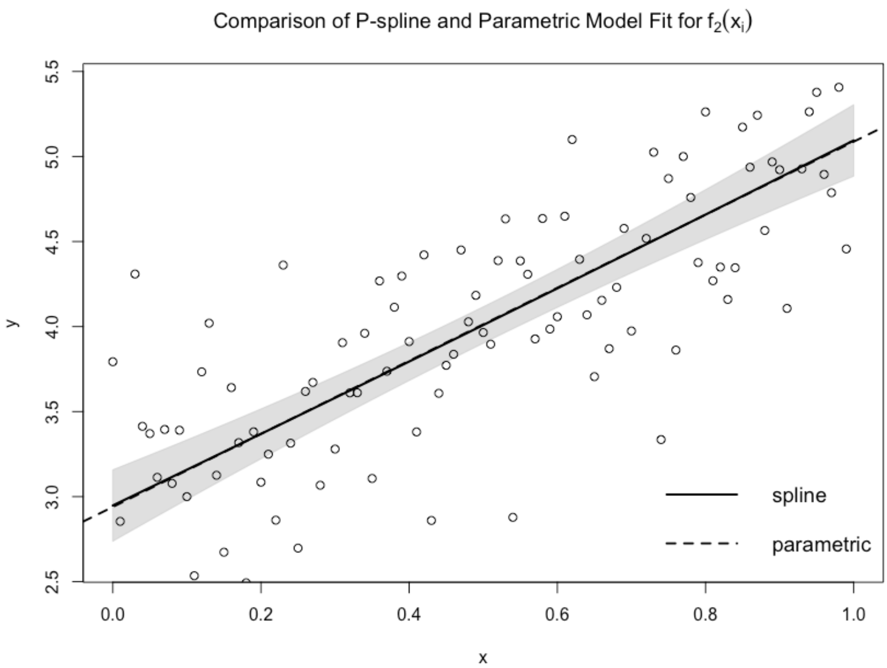


Figure 14. Comparison of P-spline and parametric model fit for $f_2(x_i)$

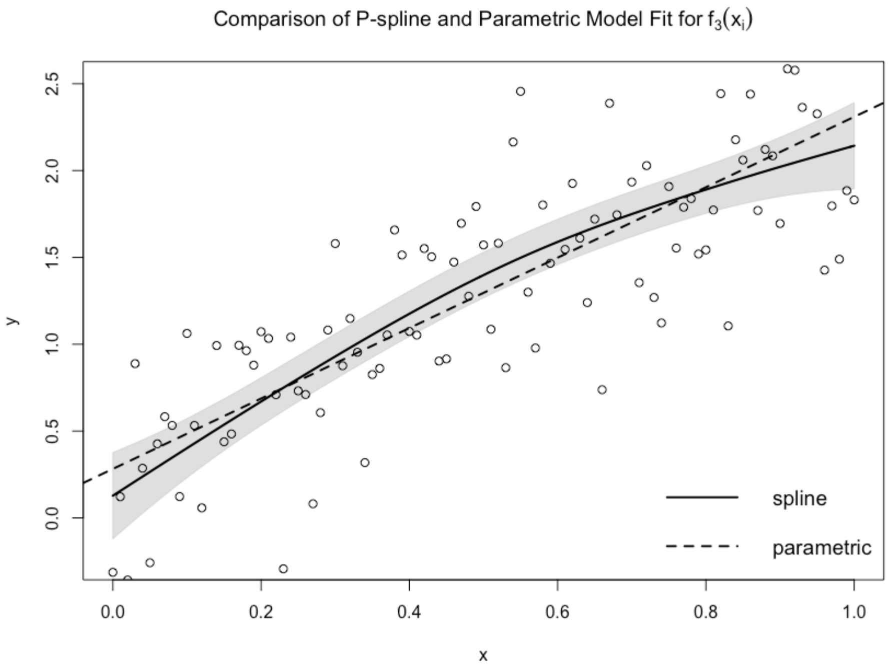


Figure 15. Comparison of P-spline and parametric model fit for $f_3(x_i)$

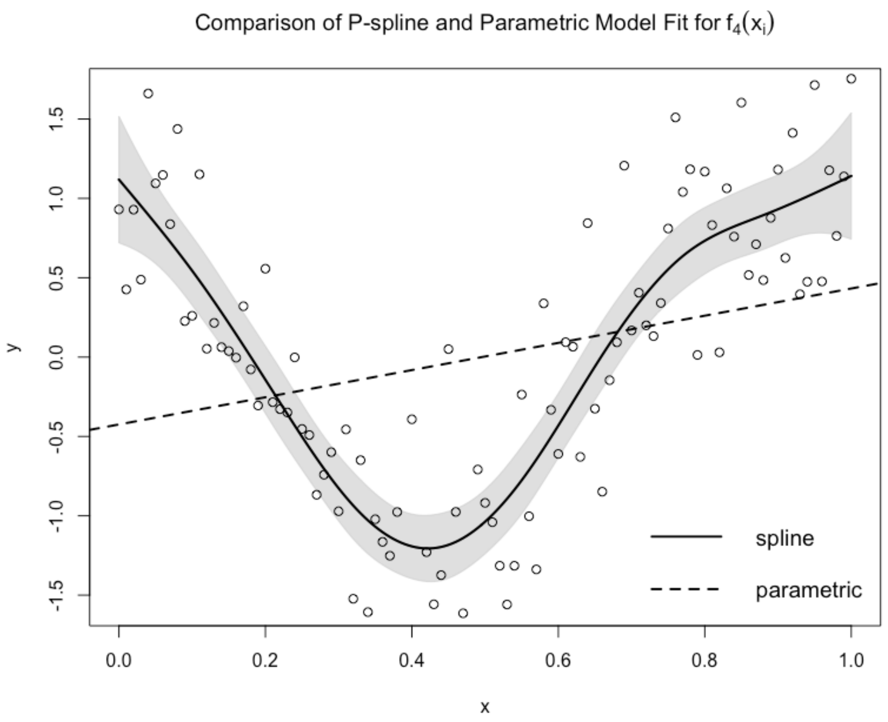


Figure 16. Comparison of P-spline and parametric model fit for $f_4(x_i)$

Likewise, for $f_3(x_i)$, the RMSE value of the P-spline is 0.0582, whereas the RMSE value of the parametric model (linear fit) is 0.1136. The RMSE value between the P-spline and the parametric model fit is 0.0822. The results indicate that while the parametric fit was within the Bayesian CI as shown in Figure 16, the nonparametric model is a better fit of the given data.

Differing from the first two data sets, for $f_4(x_i)$, the RMSE value of the P-spline is 0.1228, whereas the RMSE value of the parametric model (linear fit) is 0.6832. Also, the RMSE value between P-spline and the parametric model fit is 0.7579. Because the value of P-spline is far smaller than that of the parametric model, it is shown that the nonparametric model should be preferred in fitting the given data, consistent to Figure 16. The results from the functional case studies once again reveals that nonparametric model should be preferred in fitting the data that fluctuates with the values of x. The RMSE values for P-spline and parametric model fits for each data set are summarized in Table 5.

Table 5. RMSE value comparison from simulation

Function	RMSE Between P-spline Fit and True Function	Between Parametric Model Fit and True Function	Between P-spline Fit and Parametric Model Fit
$f_2(x_i)$	0.0453	**0.0451**	0.0038
$f_3(x_i)$	**0.0582**	0.1136	0.0822
$f_4(x_i)$	**0.1228**	0.6832	0.7579

The approximate significance of the parametric and nonparametric effects for each spline fit from simulations can be summarized as in Tables 6 and 7, respectively.

Table 6. Approx. significance of parametric effects from P-spline fits (simulation)

Function	Parameter	Estimate	Std. Error	t-value	Pr(>$\lvert t \rvert$)
$f_2(x_i)$	(Intercept)	4.013	0.0519	77.29	0***
	x	2.147	0.1793	11.97	0***
$f_3(x_i)$	(Intercept)	1.297	0.0450	28.82	0***
	x	2.015	17.847	11.29	0***
$f_4(x_i)$	(Intercept)	-0.008	0.0467	-0.16	0.8715
	x	0.003	0.2870	0.08	0.9369

*** p<0.001, ** p<0.01, * p<0.05, . p<0.1

Table 7. Approx. significance of nonparam. effects from P-spline fits (simulation)

Function	Parameter	DF	Sums of Squares	Mean Squares	F-value	Pr(>F)
$f_2(x_i)$	$f(x)$	0.082	0.0239	0.2908	1.068	0.304
	Residuals	98.918	26.9359	0.2723	-	-
$f_3(x_i)$	$f(x)$	1.053	0.9934	0.9431	4.608	0.0324*
	Residuals	97.947	20.0476	0.2047	-	-

$f_4(x_i)$	$f(x)$	4.105	59.64	14.5292	65.93	0***
	Residuals	94.895	20.91	0.2204	-	-

*** p<0.001, ** p<0.01, * p<0.05, . p<0.1

Note that the model for $f_2(x_i)$ has a coefficient of determination, R^2, of 0.5955 with 98.92 DF, and the model for $f_3(x_i)$, 0.6461 with 97.94 DF. Lastly, the model for $f_4(x_i)$ has a coefficient of determination of 0.7641 with 94.90 DF.

4.3 Comparison of Fit by Parameter Selection Method

In this section, the six smoothing parameter selection methods that were discussed in the previous chapter (OCV, GCV, AIC, BIC, ML, REML), as well as the generalized approximate cross-validation criterion (GACV) and absolute cross validation (ACV) selection methods, were implemented for comparison.

The randomly generated data created based on a noise of the function in Eq. (36) from the previous section is used. The comparison of fit is shown in Figure 17 below. To evaluate the performance of the different tuning method, the root mean squared error (RMSE) of the estimator were calculated. The fitted results and RMSE of each smoothing parameter selection method are summarized in Table 8.

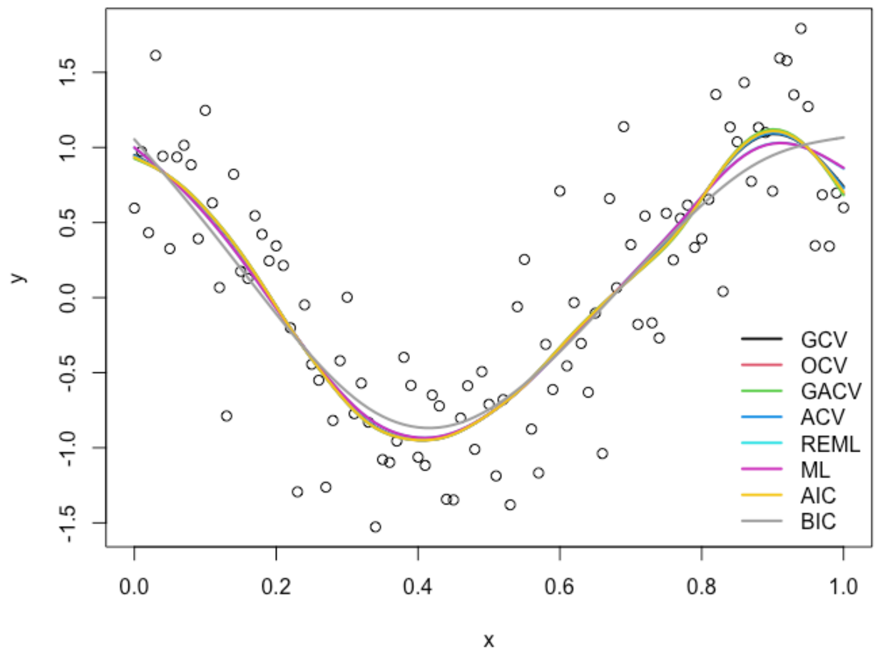


Figure 17. Comparison of fit by smoothing parameter selection method

Table 8. Summary of results and RMSE of smoothing parameter selection method

Selection Method	Criterion Value	Smoothing Parameter	Equivalent Degrees of Freedom (DF)	Penalized Criterion (RSS)	RMSE
OCV	0.2198	0.2368	8.2046	18.8486	0.0925
GCV	0.2210	0.2593	7.7839	19.0104	0.0850
GACV	-0.2353	0.2330	8.2830	18.8240	0.0938
ACV	-0.2376	0.2540	7.8854	18.9696	0.0867
ML	-70.0039	0.3143	6.7216	19.5098	0.0715

REML	-72.4965	0.3124	6.7577	19.4907	0.0718
AIC	133.4843	0.2929	8.0920	18.8900	0.0904
BIC	149.6579	0.4059	5.0999	20.6140	0.0873

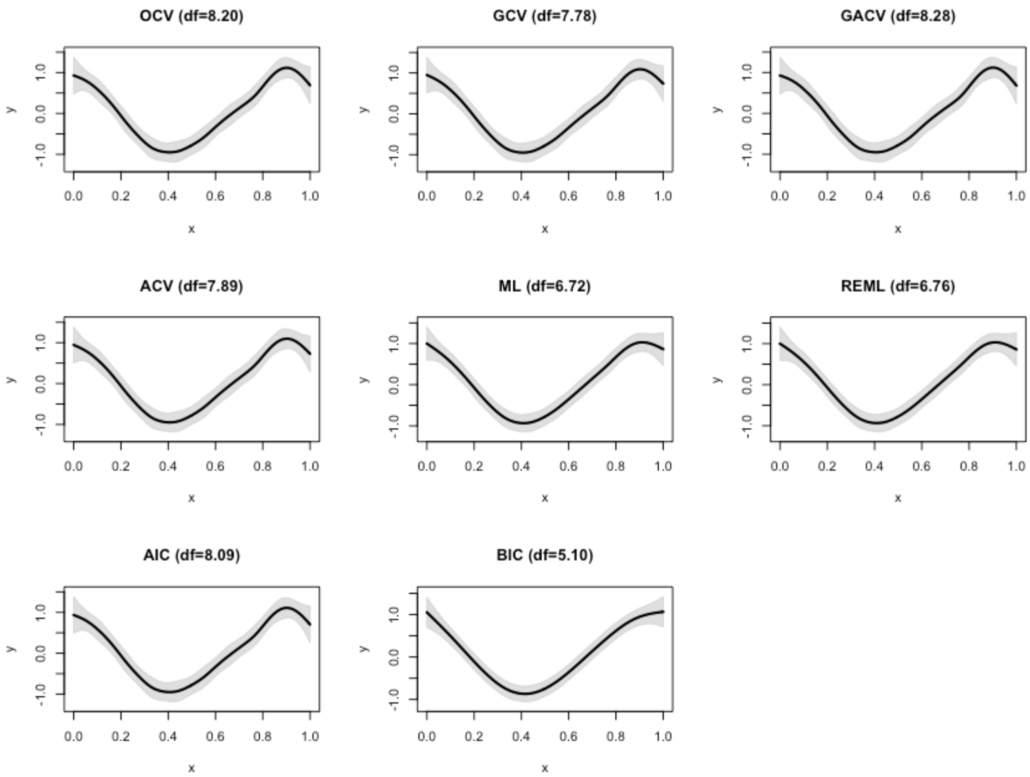


Figure 18. Simulation coverage results by smoothing parameter selection method

From the simulation results presented in Table 8, it can be noted that the parameter selection methods based on the maximum likelihood (ML and REML) tend to produce RMSE values that are similar or smaller than those produced by the methods based on cross-validation (OCV, GCV, GACV, ACV) and information theory (AIC, BIC). Then, the coverage of the 95% Bayesian confidence interval (CI) for each smoothing parameter selection method was calculated using the Eq. (44), where $I\{\cdot\}$ denotes an indicator function. Figure 18 shows the estimated functional relationship as well as the simulated coverage of the 95% Bayesian CI for each smoothing parameter selection method.

The approximate significance of the parametric effects for each spline fit from above simulation can be summarized as in Table 9.

Table 9. Approx. significance of the parametric effects by parameter selection

Selection Method	Parameter	Estimate	Std. Error	t-value	Pr(>$\|t\|$)
OCV	(Intercept)	0.0524	0.0449	1.1672	0.2461
	x	-0.2391	0.3209	-0.7451	0.4581
GCV	(Intercept)	0.0520	0.0450	1.1570	0.2502
	x	-0.2099	0.3136	-0.6694	0.5049
GACV	(Intercept)	0.0524	0.0449	1.1688	0.2455
	x	-0.2438	0.3220	-0.7572	0.4508
ACV	(Intercept)	0.0521	0.0450	1.1595	0.2492
	x	-0.2169	0.3154	-0.6879	0.4933
ML	(Intercept)	0.0511	0.0449	1.1394	0.2574

	x	-0.1361	0.2903	-0.4689	0.6402
REML	(Intercept)	0.0511	0.0453	1.1300	0.2615
	x	-0.1387	0.2938	-0.4720	0.6380
AIC	(Intercept)	0.0523	0.0449	1.1640	0.2472
	x	-0.2312	0.3189	-0.7250	0.4703
BIC	(Intercept)	0.0498	0.0462	1.0788	0.2834
	x	0.0126	0.2558	0.0493	0.9608

*** p<0.001, ** p<0.01, * p<0.05, . p<0.1

Likewise, the approximate significance of the nonparametric effects for each spline fit from simulation can be summarized as in Table 10. Note that by dividing the sum of squares by the degrees of freedom, the mean squares is obtained.

Table 10. Approx. significance of the nonparametric effects by parameter selection

Selection Method	Parameter	DF	Sums of Squares	Mean Squares	F-value	Pr($>F$)
OCV	$f(x)$	6.205	45.44	7.323	36.05	0***
	Residuals	92.795	18.85	0.203	-	-
GCV	$f(x)$	5.784	45.19	7.812	38.31	0***
	Residuals	93.216	19.01	0.204	-	-

GACV	$f(x)$	6.273	45.48	7.250	37.71	0***
	Residuals	92.727	18.82	0.203	-	-
ACV	$f(x)$	5.885	45.25	7.688	37.74	0***
	Residuals	93.115	18.97	0.204	-	-
ML	$f(x)$	4.722	44.41	9.406	46.33	0***
	Residuals	94.278	19.14	0.203	-	-
REML	$f(x)$	4.758	44.44	9.341	45.17	0***
	Residuals	94.242	19.49	0.207	-	-
AIC	$f(x)$	6.092	45.37	7.448	36.63	0***
	Residuals	92.908	18.89	0.203	-	-
BIC	$f(x)$	3.100	42.22	13.621	63.37	0***
	Residuals	95.900	20.61	0.215	-	-

*** $p<0.001$, ** $p<0.01$, * $p<0.05$, . $p<0.1$

Chapter 5. Empirical Study

The empirical application of functions with time-varying parameters is discussed in this chapter. Among the different types of consumer services, platform services have the characteristics that the switching rate (churn rate) from a predominantly used service is low due to the psychological switching costs incurred by the lock-in effect. Once consumers choose a service, they continue to invest in using that particular product rather than replacing the product they are already using, which is a potential barrier to entry for potential competitors (Klemperer, 1987, 1989; Shapiro et al., 1998). However, users also have the option of using multiple services, making it easy to find a substitute that provides a similar service, as the obvious difference between the various services tends to be difficult to discern. Since users, once they choose a service, use it repeatedly, the churn rate for ride-hailing platforms may increase if user satisfaction cannot be maintained over time. Yet, existing studies on demand-side adoption behavior of platform services do not adequately reflect the time-varying effects of continued use of the service; they utilize static utility models to explain consumer behavior with these platform services.

E-hailing ride services (ERS), also known as ride-hailing and ride-sourcing, which are on-demand services that connect car owners and passengers via smartphones, have become popular in recent years mainly because of their high-quality passenger service that provides an efficient and convenient mechanism to match supply and demand between passengers and drivers in real time through the platform (Wang, 2019; Yan et al., 2020). ERS, as one of the most representative platform services, clearly necessitates the use of a dynamic utility model in analyzing consumers' usage behaviors. As shown in Figure 19, consumers make various decisions when using ERS, such as accessing the application, requesting a ride, boarding, paying, and exiting the service. At the end of each use, they evaluate the service directly or indirectly; direct devaluation refers to leaving a rating, and indirect evaluation refers to forming an "image" or "expectation" about the service in their mind. After the initial use, the consumer decides whether to reuse the service. As such, it can be seen that considering the "repeated use" characteristic is important when evaluating consumer behavior on ERS. Therefore, this chapter aims to demonstrate the inclusion of time-varying effects of covariates in explaining the use of ERS by using models based on P-splines, a semiparametric approach.

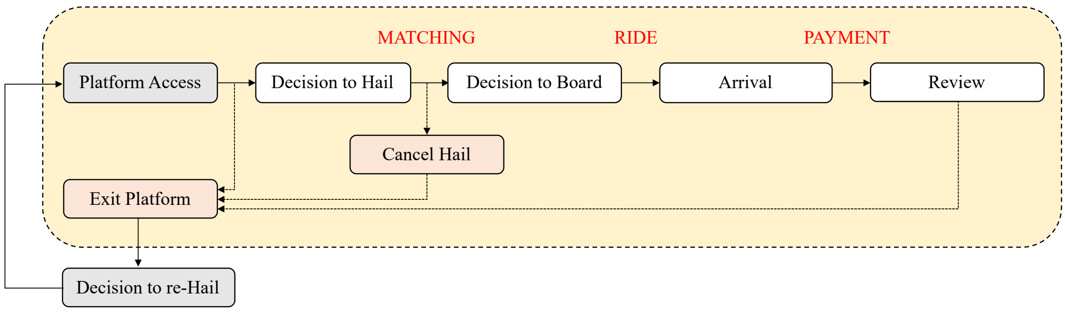


Figure 19. Passenger decision patterns on ERS

5.1 Research Background

In the taxi and travel industry, the emergence of platforms such as Uber and Lyft in the

US, Ola in India, Grab in Southeast Asia, and DiDi Chuxing in China has provided passengers and drivers a centralized and automated matching and pricing system in a two-sided marketplace, experiencing explosive growth and reshaping the urban transportation dynamics (Yan et al., 2020). Driven by the rise in trend of on-demand transportation services, creation of employment opportunities, low rate of car ownership among millennials, and global interest in the reduction of carbon emissions, the proportion of taxis and cars operated by ride-hailing companies is estimated to reach 26% by 2030 (Ferguson, 1997; Kelley, 2007; Caulfield, 2009; Chan and Shaheen, 2012; Morgan Stanley, 2016). For example, Uber, a representative transportation network company (TNC) that kickstarted the evolution of the taxi market in the early 2010s, reported that it operates in 65 countries and 700 cities, completing 4.98 billion rides with 93 million users in 2020 alone (Statistica, 2022). Grab, a leading ride-hailing platform in Southeast Asia that was founded in 2011, reported 187 million users taking an average of 46 million rides per day (Vaswani, 2021). Similarly, monthly active users (MAU) of Kakao T, a dominant ERS platform service used in South Korea, reached 10.16 million, followed by UT (860,000 MAU), a joint service of SKT's mobility subsidiary and UBER, TADA (90,000 MAU) and Macaron Taxi (30,000) of KST Mobility (Kim, 2021). It is notable that more than half of South Korean citizens are the users of these services. Nonetheless, both ridesharing and ride-sharing services reflect a shift from vehicles-as-products to vehicles-as-mobility services, with ridesharing services appealing to a much larger and broader segment of the overall population (Clewlow and Mishra, 2017).

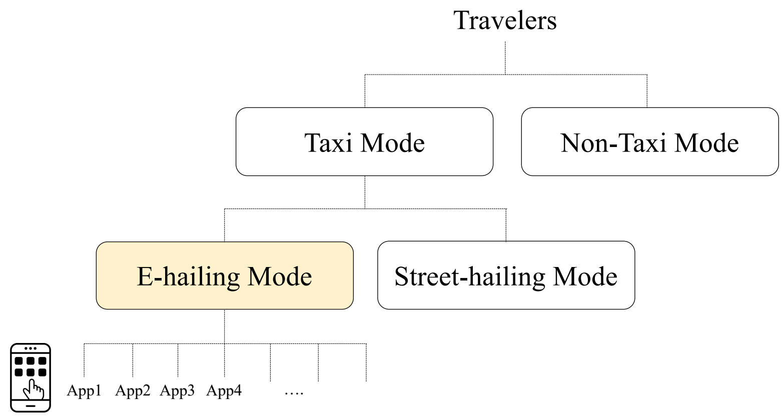


Figure 20. Consumer's choice of traveling mode and multi-homing opportunities

As the power of the platform lies in the algorithm that refines the matching based on the vast amount of data formed in the two-sided market of producer/supplier and consumer/user, a key feature of ERS is that the algorithm is a human in the task of matching available drivers with incoming requests, replacing the dispatcher. Such platforms employ two representatively different matching systems. The "inform" system allows drivers to select a destination with multiple drivers within a given area receiving the trip detail, and the first to respond is selected and assigned to the customer (Sun et al., 2020). In contrast, the "assign" system, also called first-dispatch protocol, immediately assigns an available driver who is expected to have the shortest travel time to the customer (Gao et al., 2019; Sun et al., 2020, Wei et al., 2020).

For online ride-hailing platforms, the inform system is usually chosen, which ensures drivers' freedom of choice by allowing them to control their destination by responding only to associated customer requests. However, unlike the assign system, the customer's waiting time cannot necessarily be minimized by the platform (Dai et al., 2017). Some other platforms use a system that can be considered as a hybrid of the two, where the system requests nearby drivers in order of distance until a driver decides to accept the ride; however, the key drawback of this algorithm is that the customer's average waiting time increases due to the sequential call (Sun et al., 2020). Due to these algorithmic limitations, despite the fact that these platforms aim to meet the transportation needs of travelers through seamless and efficient mobility solutions, there are still instances where individuals are sometimes not assigned to a driver at the desired time and location. Also, even when they are successfully matched, consumer satisfaction with these rides has not increased significantly.

The main stakeholders of ERS include drivers (service providers), riders/passengers (customers), the platform, and policy makers with different goals and decisions (Ashkrof et al., 2020). Since stakeholder satisfaction with a service is a very important factor for the continuity of the service, a number of studies have recently been conducted to understand the objectives of various stakeholders of ERS and to find the equilibrium between them for the operation of sustainable transport systems. Specifically with respect to the ride-hailing platform, studies have mainly attempted to explain the service adoption from both the supply (drivers) and demand side (passenger) perspectives. The assumption has been that that while the way the platform allocates rides has a direct impact on the attractiveness of the service for both drivers and passengers, not everyone has the same degree of preference for certain features of the service and they are likely to be heterogeneous in terms of their preferences.

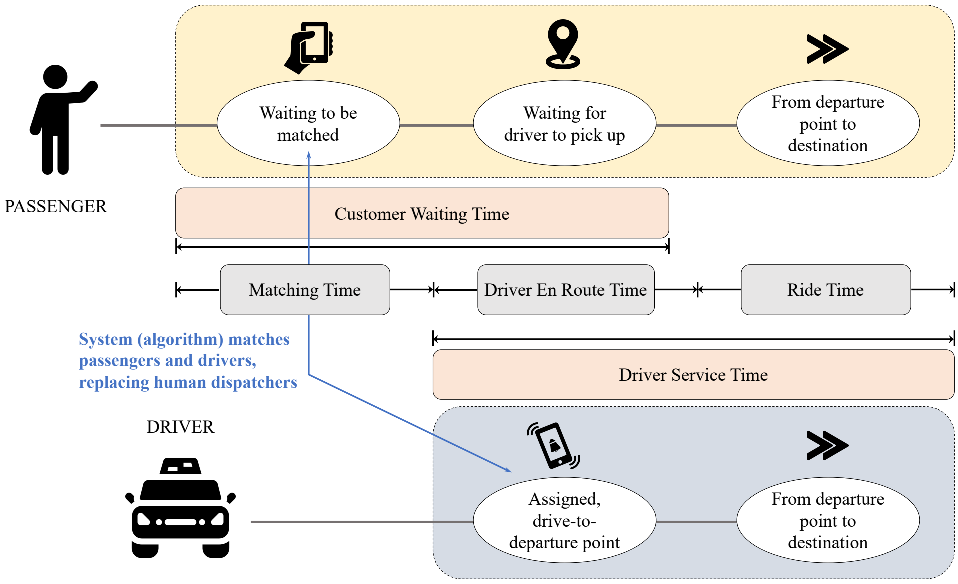


Figure 21. The role of algorithms in passenger-driver matching in ERS

Meanwhile, it has been repeatedly emphasized that not only the matching algorithm of the service must be designed to successfully meet the demand through supply, but also a better understanding of passenger heterogeneity is needed for the ride-hailing platform to make a significant contribution to satisfying consumer needs and maintaining long-term financial performance (Nguyen-Phuoc et al., 2020). Since the nature of ride-hailing platforms makes it easy for consumers to find an alternative service or choose a different mode of transportation, it is important to provide tailored services through a better understanding of passenger behavior. Since most platform services have dominant operators by geographic region, laggards seek transition from the existing profit-oriented service to a customer-oriented service in order to successfully enter and compete in the market, and in the case of dominant operators, retain their users. As previous studies have long emphasized the relationship between service quality and customer satisfaction (Rust and Oliver, 1994), the relationship between customer satisfaction and customer loyalty (Oliver, 1999; Reichheld and Sasser, 1990), is known to be the key to retaining existing passengers and attracting new passengers from other modes to ride-hailing services (Nguyen-Phuoc et al., 2020).

Therefore, this chapter aims to incorporate time-varying effects of covariates in explaining consumers' use of ride-hailing platforms by using models based on P-splines, a semiparametric approach. By doing so, accumulated experience effect rising from repeated use of service can be observed, elaborating on how the described and/or adapted 'Service Gap' influence consumers' usage behavior of the platform. While it is assumed that the customer's expectation of the service is either (1) described (specified) by the service provider or (2) formed (updated) by their own past experiences with repeated use, the chapter is expected to describe how consumer perception can update consumer expectations, and how consumers' service expectations influence their platform usage behavior over time. Then, this study identifies stream-of-time effects in consumer behaviors. The study is a meaningful addition to the literature that aims to understand platform service adoption from the demand-side perspective and provide important implications that can be used in practice when considering strategies to retain and improve the loyalty of ride-hailing platform users.

5.2 Data

The data used in this work was provided by the Macaron Taxi Company, a second largest mobility service company in South Korea. Macaron Taxi provides mobility service across South Korea (mainly Seoul), encompassing over 270,000 users and 30,000 drivers. They provide different types of services, such as immediate calling service, reservation service, calling for friends, and driving with babies/pets, etc. Considering that the immediate calling is not only the primarily used service provided by the Macaron Taxi, but also a most widely used type of service in the ride-hailing platforms, this study only considered the data from the immediate calling service for the empirical study.

As previously described in the passenger decision patterns on ride-hailing platforms, the consumer of Macaron Taxi first access the platform application when the he/she first decide to use the service to hail a taxi for the designated trip. Once the user accesses the application, one can choose whether to call the driver for him/herself or to call for friends, whether to hail now or reserve for later, etc. For this specific study, the data for immediate call for the user him/herself was categorized as the ‘immediate calling’. Then, the user is asked to choose the place of departure as well as the destination. Once all necessary information is filled out including the payment details, messages containing departure and destination locations are distributed to drivers, who can decide whether to accept the call. When a driver accepts the call, the consumer receives a message containing location and arrival time of the driver. The consumer was regarded as having successfully used the service only when they have successfully gone through the full service-usage process: matched with a driver, board the taxi, and arrived at the destination location where they make transactions for the ride. The empirical decision-making setting of Macaron Taxi's immediate call service is visualized in Figure 22 below.

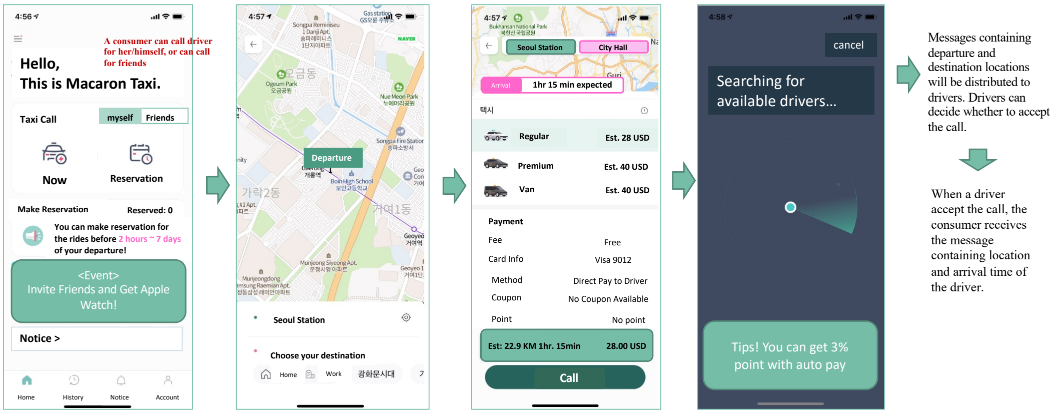


Figure 22. The empirical setting of Macaron Taxi's immediate call service

The dataset consists of individual consumer's personal information including as age, gender, and cellular phone number, as well as the detailed records of ride-hailing calls, boarding, and transactions. The data span the period from February 23 to November 14, 2020 with total of 2,502 consumers consisting of a total of 8,564 successful immediate rides. Consumer data on the usage of the service were obtained from the company's server log files, which keep track of the cookies for each consumer, call details (call request date and time), allocation details (passenger-rider matching date and time, passenger boarding date and time), and transaction details (arrival date and time, payment). The timestamp for each service-usage process has been collected and used.

After the users have successfully used the service, they were asked to evaluate the service satisfaction. Because the response was not a required part of the service, majority of the consumers did not answer their service satisfaction. The service satisfaction data from the dataset is summarized in Table 11 below.

Table 11. Service satisfaction statistics

Satisfied	Frequency	Percentage (%)
NO	39	0.46
YES	1,349	15.75
No Response	7,176	83.79

5.3 Model Specification

5.3.1 Covariates of Time and Cost

Key stakeholders of ERS include drivers (service providers), passengers/customers, platforms, and policy makers with different goals and decisions. As mentioned earlier, stakeholder satisfaction with the service is a very important factor in the success and continuity of the service. This study specifically examines the attributes that influence the user behavior of the ride-hailing platform's passengers, as user retention is a key to the success of ride-hailing platforms, where the market is highly competitive.

In the service of ride-hailing platforms, (1) the time is spent for a ride and (2) the cost of the ride are the main elements that passengers experience. Therefore, in this study, the factor of "time" was set as the main covariate among the factors that influence the usage behavior of consumers, one of the main stakeholders of the Macaron Taxi ride-hailing service. Figure 21 shows the time configuration for consumers' use of the service on this platform. The first time component is the $WaitTime$, which is the time it takes for a consumer to request a call and match a driver with a passenger, plus the time it takes for a consumer to board after the match. The second component is the $RideTime$, which is the time it takes for a passenger to board the taxi at the departure location and arrive at the destination. Therefore, the $WaitTime$ and the $RideTime$ can be represented as Eq. (49) and (50) as below using the timestamp. Then, in addition to the covariates for the time factor, the "service cost" factor was simultaneously considered as the main covariate.

$$WaitTime = (Time_{Match} - Time_{Call}) + (Time_{Board} - Time_{Match}) \quad \cdots\cdots \text{Eq. (49)}$$

$$RideTime = Time_{Arrival} - Time_{Depart} \quad \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \text{Eq. (50)}$$

5.3.2 The Interaction of Trip Distance and Travel Speed

Many people have certainly had the experience of not being able to catch a taxi at a certain time. During the commute rush hour and late at night, the demand for taxis is much higher than the supply, leaving some without matched rides. Amongst, it is known that the most problematic time is late at night. During the day, even if one cannot catch a taxi, he/she can use other means of transportation such as busses and subways. However, in the late-night hours, there are often no alternatives to get around except taxis. Without a match, he/she will not be able to go home at all. Nonetheless, endlessly waiting for a taxi or rushing home earlier than planned are not realistic alternatives either. Therefore, it is necessary to find a reasonable solution by accurately identifying when and where the excessive demand of taxis arise.

There are three types of cases in particular where there is a large excess demand for late-night taxis (Kakao Mobility Report, 2019). First, excessive demand for taxis occurred when drivers were reluctant to travel because the destined ride was too short. Meanwhile, there were also areas where the drivers did not prefer to travel, even if it was not a short-destined ride, because they were too densely populated (difficult to navigate) or there was too much traffic. Both cases lead to inefficient hourly earnings for the taxi drivers. Likewise, the drivers were hesitant to drive to the entertainment districts, such as Itaewon, Hongdae, and Gangnam Station in Seoul, even if they knew that they could easily be matched to customers. Nonetheless, even for general trips, customer’s satisfaction with the use of service may vary depending on the situation at the time the service was used, such as whether the trip was short-distanced or long-distanced, whether the service was used at a time when the road is congested or not, etc. To reflect the characteristics of the circumstances of service use, it is necessary to incorporate the variables for distance and travel speed from departure location to destination in the model.

To calculate the distance between the initial point of travel to the destination given the latitude and longitude (based on GPS), haversine formula represented in Eq. (51) can be used. Then, based on the haversine formula, Eq. (52) and (53) can be used to calculate the great-circle distance between two points – that is, the shortest distance over the earth's surface.

$$a = \sin^2\left\{{(\varphi_2 - \varphi_1)}/{2}\right\} + \cos(\varphi_1) \cdot \cos(\varphi_2) \cdot \sin^2\left\{{(\lambda_2 - \lambda_1)}/{2}\right\} \quad \cdots\cdots \text{Eq. (51)}$$

$$c = 2 \cdot \text{atan2}\left(\sqrt{a}, \sqrt{1-a}\right) \quad \cdots\cdots \text{Eq. (52)}$$

$$DirDistance = R \cdot c \quad \cdots\cdots \text{Eq. (53)}$$

In equations above, φ is the latitude and λ is longitude, where $P_{init} = (\varphi_1, \lambda_1)$ and $P_{dest} = (\varphi_2, \lambda_2)$. R represents the radius of the earth, of which the mean value is 6,371 km. It is noted that all angles are noted in radians.

Then, the average speed of travel can be calculated using the simple equation represented in Eq. (54). The $AvgSpeed$ was multiplied by 3600 to be represented in (km/hr), as all the time variables were evaluated in seconds in this empirical study.

$$AvgSpeed\ (km/hr) = {TravelDistance}/{RideTime} * 3600 \quad \cdots\cdots \text{Eq. (54)}$$

5.3.3 Formation of Consumer Expectations

Consumer expectations of the service are formed by either system specifications or usage experience. Described expectations (DE) are formed by system notifications. For example, the system notifies the consumer of (1) the expected distance from the departure location to the destination, (2) how long the trip is expected to take from the departure location to the destination, and (3) how much it is expected to cost. The expectation that the system forms through the notification are based on the system-specific algorithms. How the described expectations are formed in Macaron Taxi service is represented in Figure 23.

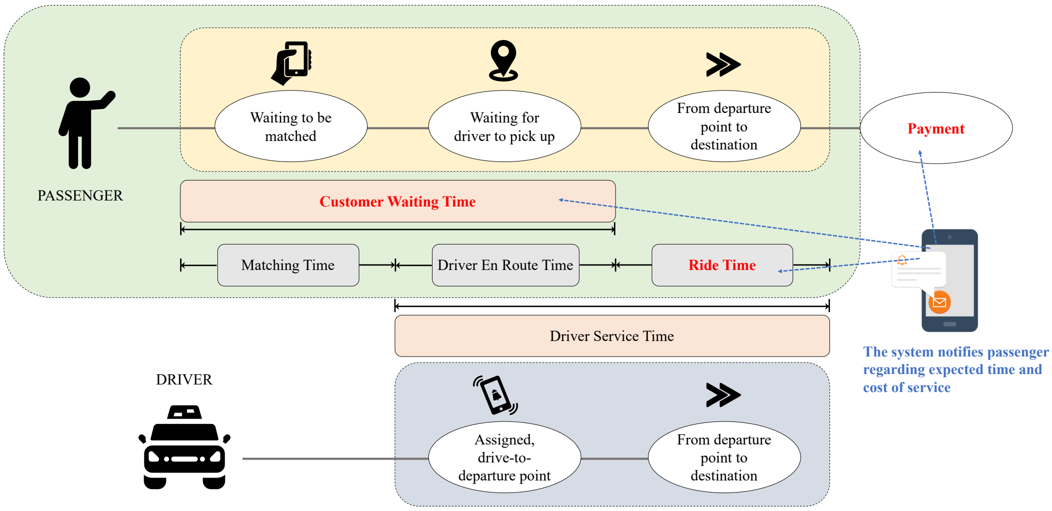


Figure 23. Formation of described expectations on covariates

Because the described expectations are not always identical to the actual delivery of the service in the ride-hailing platforms, this study incorporates a model that accounts for reference-dependent effects in explaining platform consumer behaviors. The baseline or reference effect is obtained by dividing the absolute level x_{knjt} by the gain and loss relative to the reference (Kahneman and Tversky, 2013; DellaVigna, 2009; Hess et al., 2012). If the consumer prefers an attribute with an absolute level higher than the reference point, the deterministic term in Eq. (1) can be rewritten as Eq. (55) below with respect to the reference r_{njt}.

$$V_{njt}\left(r_{knjt}\right) = I(x \geq r) \sum \beta_{nk}^{G}\left(x_{knjt} - r_{knjt}\right) + I(x < r) \sum \beta_{nk}^{L}\left|x_{knjt} - r_{knjt}\right| \quad \cdots\cdots \text{Eq. (55)}$$

In above equation, β_{nk}^{G} refers to the coefficient of gains, and β_{nk}^{L} represents the coefficient of loss. Estimated coefficients for the gain and loss domains indicate preferred and unfavorable directions relative to the reference. Prospect theory assumes an asymmetric effect of gains/losses, indicating that consumer tend to be more responsive to losses than gains (Kahneman and Tversky, 1979). However, researchers have provided conflicting empirical results about the existence of asymmetric reference price effects.

Nonetheless, as the individual repeatedly use the service, they experience gains and/or losses (difference between described expectations (system-notified time/cost) and those actually experienced (actual time/cost)) from the rides, therefore forming subjective "expectations error" regarding the service. Based on such experience, they adapt and form new expectations on gains and/or losses from the service usage (incorporating the "error"). Whereas the described expectations are readily available from our data, the experience-based updated expectations have to be computed from each consumer's purchase history. Previous research has documented the importance of incorporating purchase event feedback effects in consumer behavior analysis (Ailawadi et al., 1999). Based on Guadagni and Little (1983), the experience-based expectations on the “time” and “cost” covariates in using the service was computed recursively by exponentially smoothing past usage of service j by individual n at usage occasion $t-1$ using smoothing coefficient θ_{ref}. Such adaptive formation approach increases model fit and predictive performance and has been widely used in the marketing literature to capture covariate effects (Ailawadi et al., 1999; Kalyanaram and Winer, 1995).

In this study, we built on a widely used updating expectations framework to compute individuals’ updated reference distance, times, and cost based on their service usage history. As the individuals use the service repeatedly, they form new expectations of system errors (discrepancy) based on described expectations of service distance, times, and costs, as shown in Eq. (56), (57), and (58), respectively, below.

$$ExptErrorDist_{ijt} = \theta_{ref,dist} \cdot \left(DistGain_{ij,t-1} - DistLoss_{ij,t-1}\right) + (1-\theta_{ref,dist}) \cdot ExptErrorDist_{ij,t-1} \quad \cdots\cdots \text{Eq. (56)}$$

$$ExptErrorTime_{ijt} = \theta_{ref,time} \cdot \left(TimeGain_{ij,t-1} - TimeLoss_{ij,t-1}\right) + (1-\theta_{ref,time}) \cdot ExptErrorTime_{ij,t-1} \quad \cdots\cdots \text{Eq. (57)}$$

$$ExptErrorCost_{ijt} = \theta_{ref,cost} \cdot \left(CostGain_{ij,t-1} - CostLoss_{ij,t-1}\right) + (1-\theta_{ref,cost}) \cdot ExptErrorCost_{ij,t-1} \quad \cdots\cdots \text{Eq. (58)}$$

Here, the gains ($DistGain$, $TimeGain$, $CostGain$) and losses ($DistLoss$, $TimeLoss$, $CostLoss$) are the difference (gain/loss) between the described expectations and the actual experience of the service. Therefore, these equations represent the adapted “error” of the ride-hailing platform's described expectations. Because the users naturally adapt to the error of a system, they update their service expectations based on their experience, forming new experience-based expectations on distance, times, and costs of the service as in following Eq. (59), (60), and (61), respectively.

$$UpdatedDistE_{ijt} = DistSysE_{ijt} + ExptErrorDist_{ijt} \quad \cdots\cdots \text{Eq. (59)}$$

$$UpdatedTimeE_{ijt} = TimeSysE_{ijt} + ExptErrorTime_{ijt} \quad \cdots\cdots \text{Eq. (60)}$$

$$UpdatedCostE_{ijt} = CostSysE_{ijt} + ExptErrorCost_{ijt} \quad \cdots\cdots \text{Eq. (61)}$$

It should be noted that in above, $DistSysE_{ijt}$, $TimeSysE_{ijt}$, and $CostSysE_{ijt}$ are the reference points formed by described expectations (system-notified expectations) for that particular usage at time t. Then, the adapted error from Eq. (56) to (58) are added for each covariate to form experience-based updated expectations.

5.3.4 Estimation of Smoothing Coefficient for Error Adaption

In forming expected errors from service expectations through service experience, it is sensible to give more weight to recent observations than to observations made far in the past. Forecasts are therefore calculated using a weighted average; the weights decrease exponentially the further the observations come from. Such exponential smoothing involves the smoothing coefficient θ_{ref} with a value between 0 and 1, where the smallest weight is associated with the oldest observation. The problem that arises with this method is determining the optimal parameters to minimize the prediction error.

To determine smoothing coefficients $\theta_{ref,dist}, \theta_{ref,time}, \theta_{ref,cost}$ that minimize forecast error, the Limited-memory Broyden–Fletcher–Goldfarb–Shanno optimization algorithm (L-BFGS) is utilized using pytorch. Pytorch-LBFGS is a modular implementation of L-BFGS, a widely used popular quasi-Newton method.

Quasi-Newton methods forms an approximation to the Hessian H_k and applies a Newton-like algorithm $x_{k+1} = x_k + \theta_k H_k \nabla F(x_k)$ to solve the matrix satisfying the following secant condition.

$$H_k(x_k - x_{k-1}) = \nabla F(x_k) - \nabla(x_{k-1}) \qquad \text{Eq. (62)}$$

L-BFGS is a quasi-Newtonian optimization algorithm. Compared to the BFGS method that requires high density matrices to be stored, L-BFGS uses limited memory. Only 5-20 vectors need to be stored to implicitly approximate the matrix and the matrix-vector product is immediately constructed by two-loop recursion. In other words, this method is a method of approximating the Hessian matrix by maintaining only a few dimensional vectors instead of storing the Hessian matrix with high density.

L-BFGS constructs an approximation to the Hessian by collecting curvature pairs (s_k, y_k) in the deterministic or full-batch setting. A curvature pair is defined as the difference in consecutive gradients (i.e. $s_k = x_{k+1} - x_k \; and \; y_k = \nabla F(x_{k+1}) - \nabla(x_k)$) and iterates. In the implementation of L-BFGS algorithm in pytorch, the curvature pair is updated after an optimization step leading to x_{k+1}.

5.4 Estimation Results

In this section, the author conducts consumer behavior analysis using both (1) the generic model that do not incorporate time-varying effects and (2) functions with time-varying parameters. Then, the time-varying expectations effect (description-based vs. experience-based) are investigated. The estimation results are presented sequentially in the following sections.

5.4.1 The Generic Model

In this section linear regression models as well as a discrete choice model that do not incorporate time effects (the generic models) were used for consumer behavior analysis. To investigate the usage interval and the total number of usage of passengers in the Macaron ride-hailing platform service, negative binomial regression was used.

In the negative binomial regression, the dependent variable is an observed count that follows the negative binomial distribution. Thus, the possible values of dependent variable are the non-negative integers. It is a generalization of Poisson regression in which the limiting assumption of the Poisson model that the variance equals the mean is relaxed.

Poisson regression can be generalized by including the gamma noise variable with mean 1 and scale parameter v. In negative binomial regression, the mean of y is determined by the exposure at time t and a set of k regressors. The expressions for these variables are given in Eq. (63) below.

$$\mu_i = \exp\left(\ln(t_i) + \beta_1 x_{1i} + \beta_2 x_{2i} + \cdots + \beta_k + x_{ki}\right) \qquad \cdots\cdots\cdots\cdots\cdots\cdots\text{Eq. (63)}$$

Often $x_1 = 1$, in which case β_1 is represented as the intercept. The regression coefficients is an unknown parameter estimated in a data set. The negative binomial regression model for an observation i is then written as:

$$\Pr(\mathrm{Y} = \mathrm{y_i} | \mu_i, \alpha) = \frac{\Gamma(y_i + \alpha^{-1})}{\Gamma(\alpha^{-1})\Gamma(y_i + 1)} \left(\frac{1}{1+\alpha\mu_i}\right)^{\alpha^{-1}} \left(\frac{\alpha\mu_i}{1+\alpha\mu_i}\right)^{y_i} \qquad \cdots\cdots\cdots\cdots\cdots\text{Eq. (64)}$$

, where $\mu_i = t_i\mu$ and $\alpha = 1/v$. The parameter of μ is the mean incident rate of y per unit exposure, which can be time, space, distance, area, volume, or population size. This can be interpreted as the risk of occurrence of a new event during a given exposure.

Then, as described in the previous sections, we can assume that the service usage behavior of consumer n at occasion t depends on their demographic characteristics and the covariates of distance, time, and cost. Therefore, the service usage interval ($IntvlUse$) of consumer n at occasion t can be specified as the following Model 1 using Eq. (65). It should be noted that described expectations were used for analysis.

Model 1:

$$
\begin{aligned}
IntvlUse_{i,t} = {} & \beta_0 + \beta_1 Age_i + \beta_2 Gender_i + \beta_3 WaitTime_{i,t-1} \\
& + I(ExpDist_{i,t-1} \geq RealDist_{i,t-1})\beta_4(ExpDist_{i,t-1} - RealDist_{i,t-1}) \\
& + I(ExpDist_{i,t-1} < RealDist_{i,t-1})\beta_5|ExpDist_{i,t-1} - RealDist_{i,t-1}| \\
& + I(ExpRide_{i,t-1} \geq RealRide_{i,t-1})\beta_6(ExpRide_{i,t-1} - RealRide_{i,t-1}) \\
& + I(ExpRide_{i,t-1} < RealRide_{i,t-1})\beta_7|ExpRide_{i,t-1} - RealRide_{i,t-1}| \\
& + I(ExpCost_{i,t-1} \geq RealCost_{i,t-1})\beta_8(ExpCost_{i,t-1} - RealCost_{i,t-1}) \\
& + I(ExpCost_{i,t-1} < RealCost_{i,t-1})\beta_7|ExpCost_{i,t-1} - RealCost_{i,t-1}| \\
& + Satisfy_{i,t-1} + \epsilon_{it} \qquad \cdots\cdots\cdots \text{Eq. (65)}
\end{aligned}
$$

Our interest in using Model 1 is to estimate the influence of gains and losses of specified covariates with respect to expectations on the interval of consumers' use of service. Then, to observe the interaction effects of trip distance and travel speed as described in Section 5.3.2, Models 2 and 3 using Eq. (66) and (67) can be used as the following, respectively. The estimation results for three models are presented in Table 12.

Model 2:

$$
\begin{aligned}
IntvlUse_{i,t} = {} & \beta_0 + \beta_1 Age_i + \beta_2 Gender_i + \beta_3 WaitTime_{i,t-1} \\
& + (\beta_4 + \beta_5 Dist_{i,t-1})I(ExpDist_{i,t-1} \geq RealDist_{i,t-1})(ExpDist_{i,t-1} - RealDist_{i,t-1}) \\
& + (\beta_6 + \beta_7 Dist_{i,t-1})I(ExpDist_{i,t-1} < RealDist_{i,t-1})|ExpDist_{i,t-1} - RealDist_{i,t-1}| \\
& + (\beta_8 + \beta_9 Dist_{i,t-1})I(ExpRide_{i,t-1} \geq RealRide_{i,t-1})(ExpRide_{i,t-1} - RealRide_{i,t-1}) \\
& + (\beta_{10} + \beta_{11} Dist_{i,t-1})I(ExpRide_{i,t-1} < RealRide_{i,t-1})|ExpRide_{i,t-1} - RealRide_{i,t-1}| \\
& + (\beta_{12} + \beta_{13} Dist_{i,t-1})I(ExpCost_{i,t-1} \geq RealCost_{i,t-1})(ExpCost_{i,t-1} - RealCost_{i,t-1}) \\
& + (\beta_{13} + \beta_{14} Dist_{i,t-1})I(ExpCost_{i,t-1} < RealCost_{i,t-1})|ExpCost_{i,t-1} - RealCost_{i,t-1}| \\
& + Satisfy_{i,t-1} + \epsilon_{it} \qquad \cdots\cdots\cdots \text{Eq. (66)}
\end{aligned}
$$

Model 3:

$$
\begin{aligned}
IntvlUse_{n,t} = {} & \beta_0 + \beta_1 Age_i + \beta_2 Gender_i + \beta_3 WaitTime_{i,t-1} \\
& +(\beta_4 + \beta_5 Speed_{i,t-1}) I(ExpDist_{i,t-1} \geq RealDist_{i,t-1})(ExpDist_{i,t-1} - RealDist_{i,t-1}) \\
& +(\beta_6 + \beta_7 Speed_{i,t-1}) I(ExpDist_{i,t-1} < RealDist_{i,t-1})|ExpDist_{i,t-1} - RealDist_{i,t-1}| \\
& +(\beta_8 + \beta_9 Speed_{i,t-1}) I(ExpRide_{i,t-1} \geq RealRide_{i,t-1})(ExpRide_{i,t-1} - RealRide_{i,t-1}) \\
& +(\beta_{10} + \beta_{11} Speed_{i,t-1}) I(ExpRide_{i,t-1} < RealRide_{i,t-1})|ExpRide_{i,t-1} - RealRide_{i,t-1}| \\
& +(\beta_{12} + \beta_{13} Speed_{i,t-1}) I(ExpCost_{i,t-1} \geq RealCost_{i,t-1})(ExpCost_{i,t-1} - RealCost_{i,t-1}) \\
& +(\beta_{13} + \beta_{14} Speed_{i,t-1}) I(ExpCost_{i,t-1} < RealCost_{i,t-1})|ExpCost_{i,t-1} - RealCost_{i,t-1}| \\
& +Satisfy_{i,t-1} + \epsilon_{it} \qquad \cdots\cdots\cdots \text{Eq. (67)}
\end{aligned}
$$

The results of Model 1 show that for the ride-hailing platform, as the $WaitTime_{t-1}$ increased, the service usage interval increased. Moreover, the usage interval increased when the gained distance and cost decreased compared to expectations in the previous usage. In contrast, the gains and losses in $RideTime$ and $Satisfaction_{t-1}$ were all insignificant, indicating that they did not affect the usage interval.

It is worth noting that all of the interactive effects of age were not significant, indicating that consumer age has no effect on the influence of the covariates on the service use interval. As can be seen in Model 3, the positive coefficient of gender on $WaitTime$ and $DistanceLoss_{t-1}$ pointed to the increasing effect of wait time and distance loss with respect to expectations in the previous usage experience among female consumers. While a negative gender coefficient for $Cost\_Gain_{t-1}$ indicates an increasing effect of the cost variable among male consumers, the estimated gender coefficient for the interaction between $Cost\_Gain_{t-1}$ and $AvgSpeed_{t-1}$ was positive and significant, implying that the interactive effect of average speed decreases among female consumers. Finally, all interactive effects of travel distance were insignificant.

Table 12. Estimation results for usage interval (generic model)

VARIABLES		Model (1)	Model (2)	Model (3)
Main Effects				
Age		-4.36e-03	-4.83e-03	-4.23e-03
		(9.06e-03)	(9.19e-03)	(9.02e-03)
Gender		-1.83e-01	-2.17e-01	-1.86e-01
		(1.72e-01)	(0.170)	(1.73e-01)
WaitTime $_{t-1}$		**1.55e-04***	**1.42e-04***	**1.51e-04****
		(8.45e-05)	**(8.62e-05)**	**(8.55e-05)**
Distance	Gain $_{t-1}$	**-1.74e-01***	**-1.43e-01***	**-2.65e-01***
		(1.05e-01)	**(1.11e-01)**	**(1.17e-01)**
	Loss $_{t-1}$	3.83e-02	3.28e-02	-4.38e-02
		(9.47e-02)	(9.41e-02)	(9.41e-02)
RideTime	Gain $_{t-1}$	-1.24e-04	-4.91e-04	-1.07e-03
		(9.65e-04)	(1.46e-03)	(1.78e-03)
	Gain $_{t-1}$ × DirDistance $_{t-1}$	-	3.18e-06	-
		-	(1.43e-04)	-
	Gain $_{t-1}$ × AvgSpeed $_{t-1}$	-	-	3.40e-05
		-	-	(6.68e-05)
	Loss $_{t-1}$	3.17e-04	6.88e-04	-2.30e-03
		(9.14e-04)	(2.27e-03)	(2.20e-03)
	Loss $_{t-1}$ × DirDistance $_{t-1}$	-	-1.17e-06	-
		-	(2.02e-04)	-
	Loss $_{t-1}$ × AvgSpeed $_{t-1}$	-	-	1.02e-04
		-	-	(7.48e-05)
Cost	Gain $_{t-1}$	**-3.45e-04***	-3.42e-04	3.81e-04

	(2.09e-04)	(3.65e-04)	(5.52e-04)
Gain $_{t-1}$ × DirDistance $_{t-1}$	-	2.81e-06	-
	-	(3.80e-05)	-
Gain $_{t-1}$ × AvgSpeed $_{t-1}$	-	-	-2.96e-05
	-	-	(2.17e-05)
Loss $_{t-1}$	-2.52e-04	-2.18e-04	6.37e-04
	(1.74e-04)	(4.85e-04)	(5.89e-04)
Loss $_{t-1}$ × DirDistance $_{t-1}$	-	-1.65e-06	-
	-	(4.40e-05)	-
Loss $_{t-1}$ × AvgSpeed $_{t-1}$	-	-	-3.23e-05
	-	-	(2.07e-05)
Satisfy $_{t-1}$	-8.48e-01	**-1.03e+00***	**-1.01e+00***
	(5.36e-01)	**(5.40e-01)**	**(5.40e-01)**
Constant	**3.21e+00*** **	**3.26e+00*** **	**3.23e+00*** **
	(3.82e-01)	**(3.86e-01)**	**(3.81e-01)**
Age Interactions			
WaitTime $_{t-1}$	-3.38e-06	-3.58e-06	-3.61e-06
	(3.44e-06)	(3.46e-06)	(3.44e-06)
Distance Gain $_{t-1}$	3.10e-03	2.01e-03	4.69e-03
	(2.74e-03)	(2.96e-03)	(2.99e-03)
Loss $_{t-1}$	-2.03e-03	-1.98e-03	-2.39e-03
	(2.06e-03)	(2.04e-03)	(2.06e-03)
RideTime Gain $_{t-1}$	1.89e-05	3.23e-05	4.96e-05
	(2.14e-05)	(3.31e-05)	(3.86e-05)
Gain $_{t-1}$ × DirDistance $_{t-1}$	-	-1.09e-06	-
	-	(2.96e-06)	-
Gain $_{t-1}$ × AvgSpeed $_{t-1}$	-	-	-1.20e-06
	-	-	(1.29e-06)
Loss $_{t-1}$	8.39e-06	-1.66e-05	6.02e-05
	(2.32e-05)	(5.23e-05)	(5.54e-05)
Loss $_{t-1}$ × DirDistance $_{t-1}$	-	1.04e-06	-

		-	(4.51e-06)	-
	Loss $_{t-1}$ × AvgSpeed $_{t-1}$	-	-	-2.36e-06
		-	-	(2.02e-06)
Cost	Gain $_{t-1}$	6.34e-06	7.89e-06	-5.24e-06
		(4.93e-06)	(8.90e-06)	(1.35e-05)
	Gain $_{t-1}$ × DirDistance $_{t-1}$	-	-1.31e-07	-
		-	(8.98e-07)	-
	Gain $_{t-1}$ × AvgSpeed $_{t-1}$	-	-	4.94e-07
		-	-	(5.13e-07)
	Loss $_{t-1}$	5.36e-06	4.38e-06	-1.63e-06
		(3.76e-06)	(1.03e-05)	(1.29e-05)
	Loss $_{t-1}$ × DirDistance $_{t-1}$	-	1.16e-07	-
		-	(9.34e-07)	-
	Loss $_{t-1}$ × AvgSpeed $_{t-1}$	-	-	**8.41e-07***
		-	-	**(4.73e-07)**
Satisfy $_{t-1}$		-5.96e-03	-2.08e-03	-2.46e-03
		(1.29e-02)	(1.29e-02)	(1.29e-02)
Gender Interactions				
WaitTime $_{t-1}$		**3.28e-04***	**3.62e-04***	**3.24e-04***
		(1.90e-04)	**(1.91e-04)**	**(1.89e-04)**
Distance	Gain $_{t-1}$	5.67e-02	5.96e-02	7.39e-02
		(3.89e-02)	(5.05e-02)	(4.98e-02)
	Loss $_{t-1}$	5.45e-02	5.89e-02	**6.66e-02****
		(3.89e-02)	(4.02e-02)	**(3.98e-02)**
RideTime	Gain $_{t-1}$	-4.87e-04	-1.07e-04	5.57e-04
		(3.98e-04)	(6.36e-04)	(9.55e-04)
	Gain $_{t-1}$ × DirDistance $_{t-1}$	-	-2.69e-05	-
		-	(6.76e-05)	-
	Gain $_{t-1}$ × AvgSpeed $_{t-1}$	-	-	-3.81e-05
		-	-	(3.82e-05)
	Loss $_{t-1}$	**-5.57e-04***	2.67e-04	6.93e-04

		(3.34e-04)	(7.40e-04)	(8.42e-04)
	$Loss_{t-1}$ × $DirDistance_{t-1}$	-	-7.33e-05	-
		-	(7.22e-05)	-
	$Loss_{t-1}$ × $AvgSpeed_{t-1}$	-	-	-3.86e-05
		-	-	(2.84e-05)
Cost	$Gain_{t-1}$	6.21e-05	-1.36e-04	**-8.05e-04****
		(9.69e-05)	(1.89e-04)	**(3.39e-04)**
	$Gain_{t-1}$ × $DirDistance_{t-1}$	-	1.48e-05	-
		-	(1.57e-05)	-
	$Gain_{t-1}$ × $AvgSpeed_{t-1}$	-	-	**3.20e-05****
		-	-	**(1.27e-05)**
	$Loss_{t-1}$	4.64e-05	6.00e-05	-7.56e-05
		(4.28e-05)	(1.20e-04)	(2.03e-04)
	$Loss_{t-1}$ × $DirDistance_{t-1}$	-	-4.56e-06	-
		-	(1.38e-05)	-
	$Loss_{t-1}$ × $AvgSpeed_{t-1}$	-	-	2.61e-06
		-	-	(7.64e-06)
$Satisfy_{t-1}$		1.24e-01	1.68e-01	1.27e-01
		(2.19e-01)	(2.24e-01)	(2.19e-01)
Observations		925	925	925
/lnalpha		**8.33e-02***	7.17e-02	6.42e-02
		(4.60e-02)	(4.61e-02)	(4.62e-02)
Log-Likelihood		-3227.4408	-3321.5744	-3317.5553

*** p<0.01, ** p<0.05, * p<0.1; Standard errors are in parenthesis

Likewise, the total service use ($NbrUse$) of consumer n at occasion t can be specified as the following Model (1) using Eq. (68). Our interest in using Model 1 is to estimate the influence of gains and losses of specified covariates on the consumers' total number of uses of service. Then, to observe the interaction effects of trip distance and travel speed as described in Section 5.3.2, Models 2 and 3 using Eq. (69) and (70) can again be used for analysis. The estimation results for all three models are presented in Table 13.

Model 1:

$$
\begin{aligned}
NbrUse_{i,t} = {} & \beta_0 + \beta_1 Age_i + \beta_2 Gender_i + \beta_3 WaitTime_{i,t} \\
& + I\left(ExpDist_{i,t} \geq RealDist_{i,t}\right)\beta_4\left(ExpDist_{i,t} - RealDist_{i,t}\right) \\
& + I\left(ExpDist_{i,t} < RealDist_{i,t}\right)\beta_5\left|ExpDist_{i,t} - RealDist_{i,t}\right| \\
& + I\left(ExpRide_{i,t} \geq RealRide_{i,t}\right)\beta_6\left(ExpRide_{i,t} - RealRide_{i,t}\right) \\
& + I\left(ExpRide_{i,t} < RealRide_{i,t}\right)\beta_7\left|ExpRide_{i,t} - RealRide_{i,t}\right| \\
& + I\left(ExpCost_{i,t} \geq RealCost_{i,t}\right)\beta_8\left(ExpCost_{i,t} - RealCost_{i,t}\right) \\
& + I\left(ExpCost_{i,t} < RealCost_{i,t}\right)\beta_7\left|ExpCost_{i,t} - RealCost_{i,t}\right| \\
& + Satisfy_{i,t} + \epsilon_{it} \qquad \cdots\cdots\cdots\cdots \text{Eq. (68)}
\end{aligned}
$$

Model 2:

$$
\begin{aligned}
NbrUse_{i,t} = {} & \beta_0 + \beta_1 Age_i + \beta_2 Gender_i + \beta_3 WaitTime_{i,t} \\
& + (\beta_4 + \beta_5 Dist_{i,t}) I\left(ExpDist_{i,t} \geq RealDist_{i,t}\right)\left(ExpDist_{i,t} - RealDist_{i,t}\right) \\
& + (\beta_6 + \beta_7 Dist_{i,t}) I\left(ExpDist_{i,t} < RealDist_{i,t}\right)\left|ExpDist_{i,t} - RealDist_{i,t}\right| \\
& + (\beta_8 + \beta_9 Dist_{i,t}) I\left(ExpRide_{i,t} \geq RealRide_{i,t}\right)\left(ExpRide_{i,t} - RealRide_{i,t}\right) \\
& + (\beta_{10} + \beta_{11} Dist_{i,t}) I\left(ExpRide_{i,t} < RealRide_{i,t}\right)\left|ExpRide_{i,t} - RealRide_{i,t}\right| \\
& + (\beta_{12} + \beta_{13} Dist_{i,t}) I\left(ExpCost_{i,t} \geq RealCost_{i,t}\right)\left(ExpCost_{i,t} - RealCost_{i,t}\right) \\
& + (\beta_{13} + \beta_{14} Dist_{i,t}) I\left(ExpCost_{i,t} < RealCost_{i,t}\right)\left|ExpCost_{i,t} - RealCost_{i,t}\right| \\
& + Satisfy_{i,t} + \epsilon_{it} \qquad \cdots\cdots\cdots\cdots \text{Eq. (69)}
\end{aligned}
$$

Model 3:

$$
\begin{aligned}
NbrUse_{i,t} = {} & \beta_0 + \beta_1 Age_i + \beta_2 Gender_i + \beta_3 WaitTime_{i,t} \\
& + (\beta_4 + \beta_5 Speed_{i,t}) I\left(ExpDist_{i,t} \geq RealDist_{i,t}\right)\left(ExpDist_{i,t} - RealDist_{i,t}\right) \\
& + (\beta_6 + \beta_7 Speed_{i,t}) I\left(ExpDist_{i,t} < RealDist_{i,t}\right)\left|ExpDist_{i,t} - RealDist_{i,t}\right| \\
& + (\beta_8 + \beta_9 Speed_{i,t}) I\left(ExpRide_{i,t} \geq RealRide_{i,t}\right)\left(ExpRide_{i,t} - RealRide_{i,t}\right) \\
& + (\beta_{10} + \beta_{11} Speed_{i,t}) I\left(ExpRide_{i,t} < RealRide_{i,t}\right)\left|ExpRide_{i,t} - RealRide_{i,t}\right|
\end{aligned}
$$

$$+(\beta_{12}+\beta_{13}Speed_{i,t})I\left(ExpCost_{i,t}\geq RealCost_{i,t}\right)\left(ExpCost_{i,t}-RealCost_{i,t}\right)$$
$$+(\beta_{13}+\beta_{14}Speed_{i,t})I\left(ExpCost_{i,t}<RealCost_{i,t}\right)\left|ExpCost_{i,t}-RealCost_{i,t}\right|$$
$$+Satisfy_{i,t}+\epsilon_{it} \qquad \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \text{Eq. (70)}$$

According to the results, in the ride-hailing platform service, it was revealed that as the age of the consumers increased, the total number of service use increased. It was also found that consumers were sensitive to the cost; the higher the gain and the lower the loss compared to expectations, the higher the total number of uses, implying that "saving" cost was important. In addition, the interactive effect of distance and travel speed on $Cost\_Gain$ was negative and significant, indicating that shorter distance and lower travel speed led to a greater effect of cost gain on the number of times the service was used. Contrastingly, the main effect of gender, $WaitTime$, distance variables, and ride time variables were insignificant. One noticeable result was that the lower the satisfaction with use, the higher the total number of use.

The interactive effect of age was significant with $WaitTime$, $Cost\_Gain$, and $Satisfy$, showing that the effect of these covariates on the number of service use increased the younger the user was. The interaction effect of travel distance and average speed were both negatively significant with $Cost\_Gain$, indicating that the positive effect of cost gain on total service use was negatively affected by increasing travel distance and average speed. Furthermore, the coefficient on age was negatively significant, suggesting that the effect of increased $Cost\_Gain$ on use decreased for older consumers, while this effect amplified with increased travel distance and average speed. Gender interactions were generally not significant, but being a male resulted in a higher effect of service satisfaction on total service use, and the positive coefficient of gender on $Cost\_Loss$ pointed to the increasing effect of cost loss on female consumers.

Table 13. Estimation results for total usage (generic model)

VARIABLES		Model (1)	Model (2)	Model (3)
Main Effects				
Age		**2.98e-02*** **	**3.51e-02*** **	**3.60e-02*** **
		(8.80e-03)	**(9.17e-03)**	**(9.12e-03)**
Gender		2.38e-01	2.87e-01	1.86e-01
		(2.25e-01)	(2.21e-01)	(2.23e-01)
WaitTime		3.98e-04	3.40e-04	2.50e-04
		(3.22e-04)	(3.19e-04)	(3.23e-04)
Distance	Gain	1.90e-01	1.58e-01	1.66e-01
		(1.24e-01)	(1.22e-01)	(1.26e-01)
	Loss	5.64e-02	6.44e-02	9.28e-02
		(6.66e-02)	(7.31e-02)	(8.26e-02)
RideTime	Gain	-7.89e-04	1.35e-04	1.45e-03
		(8.11e-04)	(1.34e-03)	(1.95e-03)
	Gain × DirDistance	-	-1.25e-04	-
		-	(9.16e-05)	-
	Gain × AvgSpeed	-	-	-5.40e-05
		-	-	(4.92e-05)
	Loss	2.35e-04	1.18e-03	-5.88e-04
		(7.52e-04)	(1.24e-03)	(1.79e-03)
	Loss × DirDistance	-	-3.83e-05	-
		-	(9.17e-05)	-
	Loss × AvgSpeed	-	-	7.17e-05
		-	-	(8.51e-05)

Cost	Gain	**8.10e-04***** **(2.11e-04)**	**1.45e-03***** **(3.44e-04)**	**1.74e-03***** **(4.97e-04)**
	Gain × DirDistance	- -	**-5.92e-05***** **(1.85e-05)**	- -
	Gain × AvgSpeed	- -	- -	**-3.92e-05***** **(1.31e-05)**
	Loss	**-2.09e-04*** **(1.17e-04)**	-1.36e-04 (1.66e-04)	2.13e-06 (3.06e-04)
	Loss × DirDistance	- -	2.64e-06 (9.25e-06)	- -
	Loss × AvgSpeed	- -	- -	-1.49e-06 (8.31e-06)
Satisfy		**-7.26e-01**** **(3.18e-01)**	**-6.71e-01**** **(3.26e-01)**	**-7.22e-01**** **(3.21e-01)**
Constant		**1.37e+00***** **(3.67e-01)**	**1.10e+00***** **(3.84e-01)**	**1.16e+00***** **(3.78e-01)**
Age Interactions				
WaitTime		**-2.17e-05**** **(8.86e-06)**	**-1.85e-05**** **(8.79e-06)**	**-1.63e-05*** **(8.90e-06)**
Distance	Gain	-4.85e-03 (3.32e-03)	-5.12e-03 (3.36e-03)	-4.96e-03 (3.38e-03)
	Loss	7.78e-04 (1.44e-03)	3.30e-04 (1.65e-03)	1.37e-03 (1.95e-03)
RideTime	Gain	1.18e-05 (2.03e-05)	1.49e-05 (3.54e-05)	-3.40e-06 (4.65e-05)
	Gain × DirDistance	- -	-4.63e-07 (2.33e-06)	- -
	Gain × AvgSpeed	- -	- -	-1.74e-07 (1.18e-06)
	Loss	-1.46e-05 (1.57e-05)	-2.51e-05 (2.62e-05)	2.48e-06 (3.54e-05)

	Loss × DirDistance	-	1.28e-07	-
		-	(2.28e-06)	-
	Loss × AvgSpeed	-	-	-1.90e-06
		-	-	(1.82e-06)
Cost	Gain	**-2.01e-05*** **	**-3.08e-05*** **	**-3.88e-05*** **
		(5.18e-06)	**(8.57e-06)**	**(1.20e-05)**
	Gain × DirDistance	-	**1.24e-06*** **	-
		-	**(4.53e-07)**	-
	Gain × AvgSpeed	-	-	**8.42e-07** **
		-	-	**(3.33e-07)**
	Loss	3.73e-06	2.81e-06	1.19e-06
		(2.53e-06)	(3.60e-06)	(6.45e-06)
	Loss × DirDistance	-	-1.94e-07	-
		-	(2.01e-07)	-
	Loss × AvgSpeed	-	-	-9.36e-08
		-	-	(1.73e-07)
Satisfy		**3.00e-02*** **	**2.83e-02*** **	**2.87e-02*** **
		(7.72e-03)	**(7.94e-03)**	**(7.84e-03)**
Gender Interactions				
WaitTime		2.20e-04	1.78e-04	1.74e-04
		(1.62e-04)	(1.61e-04)	(1.62e-04)
Distance	Gain	1.52e-02	2.66e-02	1.72e-02
		(5.01e-02)	(5.40e-02)	(5.58e-02)
	Loss	**-7.98e-02** **	-5.43e-02	**-1.08e-01** **
		(3.99e-02)	(3.94e-02)	**(4.28e-02)**
RideTime	Gain	1.39e-04	-3.73e-04	-9.01e-04
		(4.06e-04)	(5.81e-04)	(8.22e-04)
	Gain × DirDistance	-	**6.83e-05***	-
		-	**(3.88e-05)**	-
	Gain × AvgSpeed	-	-	3.52e-05
		-	-	(2.28e-05)

	Loss	2.01e-04	-1.69e-04	4.35e-04
		(3.04e-04)	(4.83e-04)	(6.95e-04)
	Loss × DirDistance	-	2.11e-05	-
		-	(3.21e-05)	-
	Loss × AvgSpeed	-	-	-5.56e-06
		-	-	(3.11e-05)
Cost	Gain	2.24e-05	-1.36e-04	-3.40e-06
		(9.56e-05)	(1.41e-04)	(1.99e-04)
	Gain × DirDistance	-	3.59e-06	-
		-	(5.56e-06)	-
	Gain × AvgSpeed	-	-	-1.22e-07
		-	-	(4.80e-06)
	Loss	**5.32e-05***	3.67e-05	1.31e-05
		(3.00e-05)	(4.41e-05)	(8.96e-05)
	Loss × DirDistance	-	1.09e-06	-
		-	(3.77e-06)	-
	Loss × AvgSpeed	-	-	1.48e-06
		-	-	(3.17e-06)
Satisfy		**-4.89e-01****	**-4.66e-01****	**-4.18e-01****
		(2.07e-01)	**(2.04e-01)**	**(2.06e-01)**
Observations		1,388	1,388	1,388
/lnalpha		**3.10e-01*****	**2.57e-01*****	**2.72e-01*****
		(3.59e-02)	**(3.64e-02)**	**(3.62e-02)**
Log-Likelihood		-5037.6187	-4994.8908	-5005.5464

*** p<0.01, ** p<0.05, * p<0.1; Standard errors are in parenthesis

Finally, an ordered logit model (OL) was used to examine service satisfaction among users of the Macaron ride-hailing platform service. OL can be applied when the dependent variable Y is categorical and has a significant order with three or more categories or levels.

The ordinal variable Y is a function of another variable Y^* that is continuous and unmeasured and has different critical points. The value Y_i of the observed variable depends on whether or not a certain threshold has been crossed, as shown by the formula:

$$Y_i = \begin{cases} 1 & if\ Y_i^* \leq k_1 \\ j & if\ k_i \leq Y_i^* \leq k_{i-1} \\ M & if\ Y_i^* \geq k_{M-1} \end{cases} \quad \cdots\cdots \text{Eq. (71)}$$

The continuous latent variable Y^* is equal to Eq. (72), where the stochastic term ϵ_i is normally distributed. The vector of β parameters is estimated by the Maximum Likelihood method.

$$Y_i = \sum_{k=1}^{K} \beta_k X_{ki} + \epsilon_i \quad \cdots\cdots \text{Eq. (72)}$$

The probability of each categorical outcome j is equal to the probability that the linear function estimated with the random error lies within the range of cutpoints computed for the outcome as follows:

$$\Pr\,(Y_i = j) = \Pr(k_{i-1} < \beta_1 x_{1i} + \cdots + \beta_k x_{ki} + \epsilon_i \leq k_i) \quad \cdots\cdots \text{Eq. (73)}$$

In our study, the OL was adopted to identify factors that influence consumer's satisfaction of the service, which is identified as 1=satisfied, 0=indifferent, and -1 = not satisfied. No response was categorized as consumers being indifferent.

Again, we can assume that the service satisfaction of consumer n at occasion t depends on their demographic characteristics and the covariates of distance, time, and cost. In addition, consumer's satisfaction in the previous service use was incorporated. Therefore, the service satisfaction of consumer n at occasion t can be specified as the following:

Model 1:

$$
\begin{aligned}
Satisfy_{i,t} = {} & \beta_0 + \beta_1 Age_i + \beta_2 Gender_i + \beta_3 WaitTime_{i,t} \\
& + I\left(ExpDist_{i,t} \geq RealDist_{i,t}\right)\beta_4\left(ExpDist_{i,t} - RealDist_{i,t}\right) \\
& + I\left(ExpDist_{i,t} < RealDist_{i,t}\right)\beta_5\left|ExpDist_{i,t} - RealDist_{i,t}\right| \\
& + I\left(ExpRide_{i,t} \geq RealRide_{i,t}\right)\beta_6\left(ExpRide_{i,t} - RealRide_{i,t}\right) \\
& + I\left(ExpRide_{i,t} < RealRide_{i,t}\right)\beta_7\left|ExpRide_{i,t} - RealRide_{i,t}\right| \\
& + I\left(ExpCost_{i,t} \geq RealCost_{i,t}\right)\beta_8\left(ExpCost_{i,t} - RealCost_{i,t}\right) \\
& + I\left(ExpCost_{i,t} < RealCost_{i,t}\right)\beta_7\left|ExpCost_{n,t} - RealCost_{i,t}\right| \\
& + \epsilon_{nt} \qquad \cdots\cdots \text{Eq. (74)}
\end{aligned}
$$

Model 2:

$$
\begin{aligned}
Satisfy_{i.t} = {} & \beta_0 + \beta_1 Age_i + \beta_2 Gender_i + \beta_3 WaitTime_{i,t} \\
& + (\beta_4 + \beta_5 Dist_{i,t}) I\left(ExpDist_{i,t} \geq RealDist_{i,t}\right)\left(ExpDist_{i,t} - RealDist_{i,t}\right) \\
& + (\beta_6 + \beta_7 Dist_{i,t}) I\left(ExpDist_{i,t} < RealDist_{i,t}\right)\left|ExpDist_{i,t} - RealDist_{i,t}\right| \\
& + (\beta_8 + \beta_9 Dist_{i,t}) I\left(ExpRide_{i,t} \geq RealRide_{i,t}\right)\left(ExpRide_{i,t} - RealRide_{i,t}\right) \\
& + (\beta_{10} + \beta_{11} Dist_{i,t}) I\left(ExpRide_{i,t} < RealRide_{i,t}\right)\left|ExpRide_{i,t} - RealRide_{i,t}\right| \\
& + (\beta_{12} + \beta_{13} Dist_{i,t}) I\left(ExpCost_{i,t} \geq RealCost_{i,t}\right)\left(ExpCost_{i,t} - RealCost_{i,t}\right) \\
& + (\beta_{13} + \beta_{14} Dist_{i,t}) I\left(ExpCost_{i,t} < RealCost_{i,t}\right)\left|ExpCost_{i,t} - RealCost_{i,t}\right| \\
& + \epsilon_{it} \qquad \cdots\cdots \text{Eq. (75)}
\end{aligned}
$$

Model 3:

$$
\begin{aligned}
Satisfy_{i.t} = {} & \beta_0 + \beta_1 Age_i + \beta_2 Gender_i + \beta_3 WaitTime_{i,t} \\
& + (\beta_4 + \beta_5 Speed_{i,t}) I\left(ExpDist_{i,t} \geq RealDist_{i,t}\right)\left(ExpDist_{i,t} - RealDist_{i,t}\right) \\
& + (\beta_6 + \beta_7 Speed_{i,t}) I\left(ExpDist_{i,t} < RealDist_{i,t}\right)\left|ExpDist_{i,t} - RealDist_{i,t}\right|
\end{aligned}
$$

$$+(\beta_8 + \beta_9 Speed_{i,t}) I(ExpRide_{i,t} \geq RealRide_{i,t})(ExpRide_{i,t} - RealRide_{i,t})$$
$$+(\beta_{10} + \beta_{11} Speed_{i,t}) I(ExpRide_{i,t} < RealRide_{i,t})|ExpRide_{i,t} - RealRide_{i,t}|$$
$$+(\beta_{12} + \beta_{13} Speed_{i,t}) I(ExpCost_{i,t} \geq RealCost_{i,t})(ExpCost_{i,t} - RealCost_{i,t})$$
$$+(\beta_{13} + \beta_{14} Speed_{i,t}) I(ExpCost_{i,t} < RealCost_{i,t})|ExpCost_{i,t} - RealCost_{i,t}|$$
$$+\epsilon_{it} \quad \cdots\cdots \text{Eq. (76)}$$

Our interest in using Model 1 is to estimate the influence of gains and losses of specified covariates on the consumers' satisfaction of service. Then, to observe the interaction effects of trip distance and travel speed as described in Section 5.3.2, Models 2 and 3 using Eq. (75) and (76) can be used as the following, respectively. The estimation results for three models are presented in Table 14.

According to the results, the younger the user, the higher the user satisfaction. The satisfaction was also higher for male than for female It was also revealed that the consumer's satisfaction in the previous use of service positively influences the service satisfaction of the current use. All other covariates with regards to time and cost were insignificant, including the interaction effects of travel distance and average speed.

Table 14. Estimation results for service satisfaction (generic model)

VARIABLES	Model (1)	Model (2)	Model (3)
Main Effects			
Age	**-1.43e-02*** **	**-1.43e-02*** **	**-1.29e-02** **
	(5.27e-03)	**(5.34e-03)**	**(5.34e-03)**
Gender	**-3.34e-01*** **	**-4.07e-01*** **	**-4.23e-01*** **
	(9.52e-02)	**(1.01e-02)**	**(1.02e-01)**
WaitTime	-2.88e-04	-3.67e-04	-3.45e-04

		(3.20e-04)	(3.22e-04)	(3.21e-04)
Distance	Gain	-2.90e-02	-1.18e-02	-5.30e-02
		(6.57e-02)	(6.90e-02)	(6.97e-02)
	Loss	4.19e-02	6.28e-02	7.40e-02
		(5.21e-02)	(4.76e-02)	(5.65e-02)
RideTime	Gain	-3.07e-04	-1.25e-03	-9.36e-04
		(5.76e-04)	(7.78e-04)	(8.74e-04)
	Gain × DirDistance	-	9.57e-05*	-
		-	(5.28e-05)	-
	Gain × AvgSpeed	-	-	2.07e-05
		-	-	(2.44e-05)
	Loss	-6.01e-04	-1.25e-03	-7.66e-05
		(5.08e-04)	(8.71e-04)	(7.97e-04)
	Loss × DirDistance	-	6.34e-05	-
		-	(5.61e-05)	-
	Loss × AvgSpeed	-	-	-2.80e-05
		-	-	(4.30e-05)
Cost	Gain	1.28e-04	1.82e-04	3.44e-04
		(1.52e-04)	(2.43e-04)	(3.04e-04)
	Gain × DirDistance	-	-8.77e-06	-
		-	(1.35e-05)	-
	Gain × AvgSpeed	-	-	-5.96e-06
		-	-	(7.94e-06)
	Loss	1.50e-05	-6.63e-05	-9.89e-05
		(9.37e-05)	(1.52e-04)	(1.69e-04)
	Loss × DirDistance	-	1.62e-06	-
		-	(4.79e-06)	-
	Loss × AvgSpeed	-	-	3.69e-06
		-	-	(3.39e-06)
$Satisfy_{t-1}$		**1.60e+00*** **	**1.61e+00*** **	**1.61e+00*** **
		(7.63e-02)	**(7.66e-02)**	**(7.66e-02)**

Age Interactions				
WaitTime		2.41e-07	1.70e-06	1.57e-06
		(8.14e-06)	(8.20e-06)	(8.17e-06)
Distance	Gain	4.21e-04	9.83e-05	1.14e-03
		(1.68e-03)	(1.79e-03)	(1.81e-03)
	Loss	-8.61e-04	-9.52e-04	-1.33e-03
		(1.30e-03)	(1.16e-03)	(1.40e-03)
RideTime	Gain	1.09e-05	2.25e-05	9.14e-06
		(1.45e-05)	(1.93e-05)	(2.25e-05)
	Gain × DirDistance	-	-1.80e-06	-
		-	(1.25e-06)	-
	Gain × AvgSpeed	-	-	-2.16e-07
		-	-	(6.28e-07)
	Loss	1.34e-05	2.89e-05	4.12e-06
		(1.16e-05)	(1.97e-05)	(1.84e-05)
	Loss × DirDistance	-	-1.68e-06	-
		-	(1.32e-06)	-
	Loss × AvgSpeed	-	-	3.33e-07
		-	-	(9.67e-07)
Cost	Gain	-5.94e-07	-2.09e-06	-5.41e-06
		(3.56e-06)	(5.38e-06)	(6.54e-06)
	Gain × DirDistance	-	2.66e-07	-
		-	(3.09e-07)	-
	Gain × AvgSpeed	-	-	1.63e-07
		-	-	(1.78e-07)
	Loss	-1.51e-07	1.67e-06	2.61e-06
		(2.04e-06)	(3.32e-06)	(3.66e-06)
	Loss × DirDistance	-	-3.87e-08	-
		-	(1.05e-07)	-
	Loss × AvgSpeed	-	-	-9.32e-08
		-	-	(7.30e-08)

Gender Interactions				
WaitTime		-4.82e-05	-3.55e-05	-4.47e-05
		(1.59e-04)	(1.60e-04)	(1.60e-04)
Distance	Gain	1.33e-02	1.31e-02	1.32e-02
		(1.69e-02)	(1.70e-02)	(1.70e-02)
	Loss	2.48e-03	1.93e-03	1.70e-03
		(6.30e-03)	(6.31e-03)	(6.31e-03)
RideTime	Gain	7.53e-04	**8.54e-04****	**1.07e-03****
		(3.32e-04)	**(3.89e-04)**	**(5.12e-04)**
	Gain × DirDistance	-	-3.83e-05	-
		-	(2.47e-05)	-
	Gain × AvgSpeed	-	-	-1.41e-05
		-	-	(1.49e-05)
	Loss	-2.40e-04	-2.80e-04	**-6.64e-04***
		(2.13e-04)	(3.27e-04)	**(3.79e-04)**
	Loss × DirDistance	-	6.25e-06	-
		-	(1.95e-05)	-
	Loss × AvgSpeed	-	-	2.62e-05
		-	-	(1.91e-05)
Cost	Gain	**-1.51e-04****	-1.52e-04	**-2.36e-04***
		(6.73e-05)	(1.08e-04)	**(1.29e-04)**
	Gain × DirDistance	-	-3.35e-06	-
		-	(6.18e-06)	-
	Gain × AvgSpeed	-	-	1.10e-07
		-	-	(3.24e-06)
	Loss	-1.74e-05	1.06e-05	5.57e-05
		(3.40e-05)	(3.86e-05)	(5.76e-05)
	Loss × DirDistance	-	-9.15e-07	-
		-	(1.11e-06)	-
	Loss × AvgSpeed	-	-	-2.24e-06
		-	-	(1.70e-06)

/cut1	**-6.06e+00*** (2.78e-01)**	**-6.11e+00*** (2.82e-01)**	**-6.06e+00*** (2.81e-01)**
/cut2	**1.26e+00*** (2.23e-01)**	**1.23e+00*** (2.28e-01)**	**1.27e+00*** (2.27e-01)**
Observations	8,564	8,564	8,564
Log-Likelihood	-3709.441	-3698.702	-3700.3472

*** p<0.01, ** p<0.05, * p<0.1; Standard errors are in parenthesis

5.4.2 Accumulated Experience Effect

In this section, consumer satisfaction with the service is classified by number of uses (i.e., first use, second use, etc.) to show that people's preferences may change over time with continued use of the service. The time-varying effects of covariates in explaining consumers' use of ride-hailing platforms are analyzed, first based on the described expectations. Then, updated expectations are used to show how users may make experience-based decisions based on direct or vicarious reinforcement they have received in the past, and to show how preference for a service is influenced by past usage experiences. While the described expectations are readily available from our data, the experience-based updated expectations need to be computed from individual consumers purchase history, as mentioned in Section 5.3.3. Finally, updated perceptions of the gain and/or loss of a service component are estimated using the Long Short-Term Memory (LSTM) model, a form of Recurrent Neural Network (RNN), to illustrate the impact of consumers' updated perceptions on service satisfaction. The time-varying effects of covariates based on described expectations, updated expectations, and updated perceptions are compared and presented. In all cases, service user preferences are captured with models based on P-splines, a semiparametric approach. In Figure 24, individuals are categorized by the total number of successful rides.

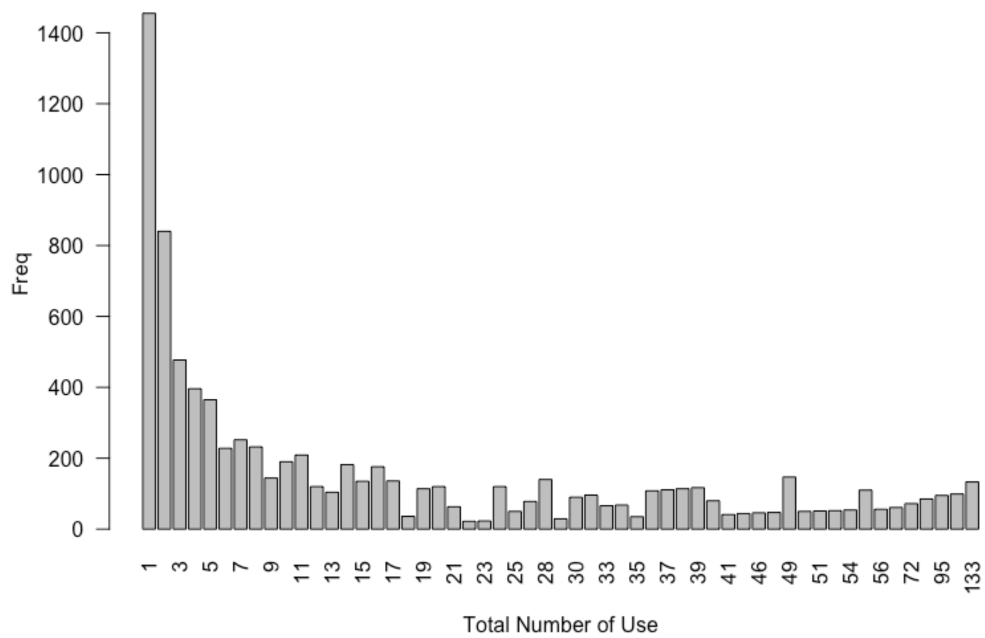


Figure 24. Total number of service use by individuals

5.4.2.1 Change of Preference with Repeated Use

First, Macaron Taxi data, which included a total of 8,564 successful immediate rides, were classified by number of uses (i.e., first use, second use, etc.) to analyze consumers' accumulated experience effects. For each ride, the number of times the individual has used the service was identified and classified. The frequency of data classified by the usage count is presented in Figure 25. The empirical model used to analyze the accumulated experience effects of the covariates on service satisfaction ($Satisfy$) of consumers of the Macaron ride-hailing platform service is shown in Equation (77) using OL regression. Again, our interest is to determine the impact of gains and losses of certain covariates, i.e., the "Service Gap", on consumers' service satisfaction and how it varies with the number of times the service is used. In the model, the seasonal effect (i.e., weekday/weekend and the time of the day) was added as a control variable. The estimation results are presented in Table 15, and the significant variables for each time of day are again summarized in Table 16 for clarity. The conceptualization of the "Service Gap" is illustrated in Figure 26.

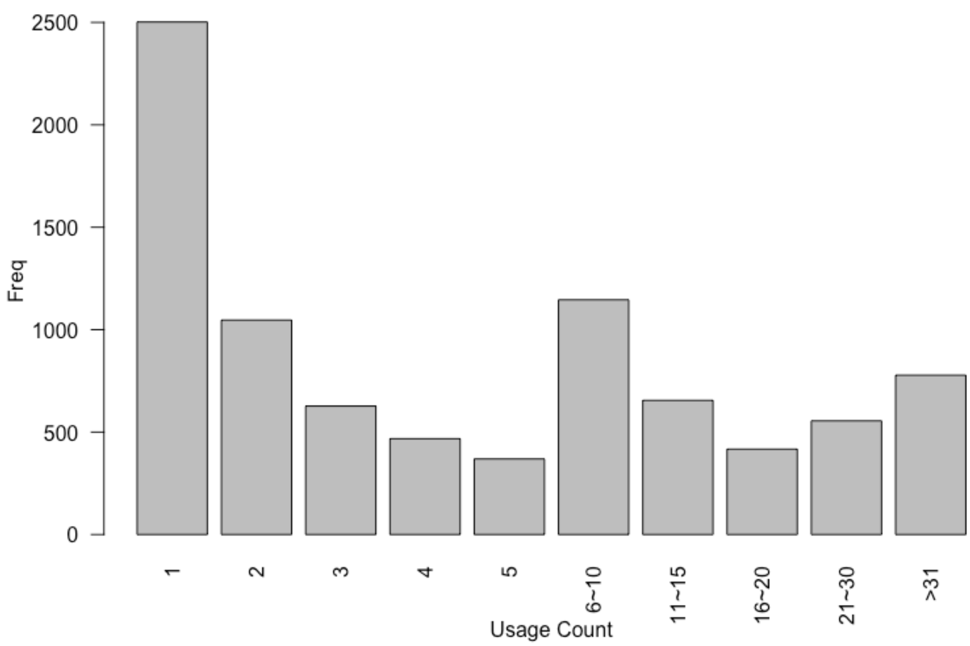


Figure 25. Frequency of data with usage count

$$
\begin{aligned}
Satisfy_{i,t} = {} & \beta_0 + \beta_1 Age_i + \beta_2 Gender_i + \beta_3 WaitTime_{i,t} + \beta_4 Weekend_{i,t} + \beta_5 Daytime_{i,t} \\
& + \beta_6 RushHour_{i,t} + \beta_7 LateNight_{i,t} + \beta_8 EarlyMorning_{i,t} \\
& + I(SystemDistE_{i,t} \geq RealDist_{i,t})\beta_9(UpdatedDistE_{i,t} - RealDist_{i,t}) \\
& + I(SystemDistE_{i,t} < RealDist_{i,t})\beta_{10}|UpdatedDistE_{i,t} - RealDist_{i,t}| \\
& + I(SystemTimeE_{i,t} \geq RealTime_{i,t})\beta_{11}(UpdatedTimeE_{i,t} - RealRide_{i,t}) \\
& + I(SystemTimeE_{i,t} < RealRide_{n,t})\beta_{12}|UpdatedTimeE_{i,t} - RealRide_{i,t}| \\
& + I(SystemCostE_{i,t} \geq RealCost_{i,t})\beta_{13}(UpdatedCostE_{i,t} - RealCost_{i,t}) \\
& + I(SystemCostE_{i,t} < RealCost_{i,t})\beta_{14}|UpdatedCostE_{i,t} - RealCost_{,t}| \\
& + \epsilon_{it} \qquad \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \text{Eq. (77)}
\end{aligned}
$$

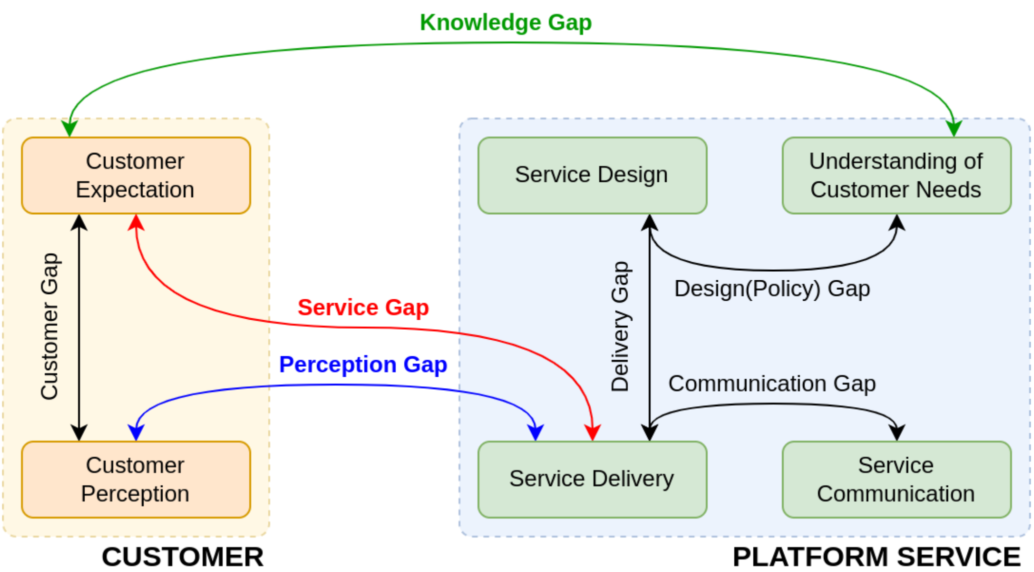


Figure 26. Conceptualization of the “Service Gap” and “Perception Gap”

Table 15. Estimation results for service satisfaction with repeated use (DE)

VARIABLES		Cnt=1	Cnt=2	Cnt=3	Cnt=4	Cnt=5	Cnt=6~10	Cnt=11~15
Age		-6.48e-03	**-2.18e-02****	-1.99e-02	-1.43e-03	-4.36e-03	-1.57e-02	**2.67e-02***
		(5.45e-03)	**(1.04e-02)**	(1.35e-02)	(1.42e-02)	(2.10e-02)	(1.04e-02)	**(1.37e-02)**
Gender		**-2.94e-01*****	**-4.44e-01****	**-6.81e-01*****	-2.77e-01	**-7.13e-01***	**-7.06e-01*****	**-5.46e-01****
		(1.04e-01)	**(1.85e-01)**	**(2.48e-01)**	(2.86e-01)	**(3.70e-01)**	**(1.94e-01)**	**(2.63e-01)**
Weekend		-2.76e-02	-1.44e-01	1.99e-01	-9.51e-02	-2.80e-01	4.43e-02	1.61e-01
		(1.11e-01)	(2.07e-01)	(2.85e-01)	(3.53e-01)	(5.33e-01)	(2.59e-01)	(3.37e-01)
Time of the	Daytime	1.11e-01	-1.32e-01	-3.79e-01	3.11e-01	1.26e-01	1.17e-01	-2.07e-01
Day	(10am-4pm)	(1.11e-01)	(1.88e-01)	(2.58e-01)	(2.94e-01)	(3.89e-01)	(2.15e-01)	(2.86e-01)
	Rush Hour	2.48e-01	-1.16e-01	-1.08e-01	5.09e-01	-3.88e-01	2.21e-02	-5.63e-02
	(7-9am, 5-8pm)	(1.54e-01)	(2.79e-01)	(3.82e-01)	(4.58e-01)	(7.15e-01)	(3.35e-01)	(4.26e-01)
	Evening (9-11pm)				(reference)			
	Late Night	2.50e-01	2.28e-01	-3.89e-01	1.46e-01	5.48e-01	2.61e-02	-6.12e-01
	(12am-3am)	(2.06e-01)	(3.56e-01)	(5.33e-01)	(5.87e-01)	(6.75e-01)	(3.87e-01)	(5.74e-01)
	Early Morning	1.43e-02	-5.94e-01	1.54e-01	6.08e-01	-14.84	8.90e-02	-1.203
	(4-6am)	(2.06e-01)	(4.75e-01)	(5.01e-01)	(6.09e-01)	(1,039)	(4.11e-01)	(8.45e-01)
WaitTime		**-3.44e-04*****	**-4.82e-04****	5.72e-05	**-6.72e-04***	**-7.72e-04***	**-4.57e-04***	2.50e-04
		(1.21e-04)	**(2.04e-04)**	(2.38e-04)	**(3.52e-04)**	**(4.67e-04)**	**(2.53e-04)**	(4.06e-04)
Distance	Gain	-1.79e-02	**-1.14e-01***	-3.36e-02	-4.32e-03	-3.46e-01	2.94e-02	4.70e-02

		(1.92e-02)	**(6.08e-02)**	(9.82e-02)	(8.93e-02)	(2.32e-01)	(6.75e-02)	(7.62e-02)
	Loss	1.22e-04	-7.58e-03	-4.58e-02	-9.70e-02	-5.83e-02	**3.85e-02***	**2.05e-01*****
		(8.95e-03)	(3.16e-02)	(3.89e-02)	(1.06e-01)	(9.57e-02)	**(1.99e-02)**	**(6.36e-02)**
RideTime	Gain	**4.43e-04****	-2.13e-04	**-1.21e-03***	-3.36e-04	-4.39e-04	-1.82e-04	-1.00e-03
		(2.10e-04)	(4.61e-04)	**(6.43e-04)**	(6.59e-04)	(1.22e-03)	(5.68e-04)	(7.43e-04)
	Loss	-2.42e-04	-3.86e-05	-4.47e-04	-3.83e-04	1.75e-03	**-6.34e-04***	**-8.30e-04***
		(1.58e-04)	(2.48e-04)	(4.56e-04)	(7.23e-04)	(1.08e-03)	**(3.67e-04)**	**(4.46e-04)**
Cost	Gain	1.55e-05	1.78e-05	1.39e-04	-5.26e-06	2.57e-04	-1.38e-04	2.81e-04
		(3.59e-05)	(1.17e-04)	(1.40e-04)	(1.76e-04)	(3.22e-04)	(1.54e-04)	(1.97e-04)
	Loss	**-4.56e-05****	**4.35e-05***	-7.82e-06	-5.55e-05	-2.46e-04	4.18e-05	**-1.89e-04****
		(1.90e-05)	**(2.27e-05)**	(4.70e-05)	(2.10e-04)	(2.68e-04)	(7.02e-05)	**(9.23e-05)**
/cut1		**-5.987*****	**-7.117*****	**-7.148*****	**-4.112*****	**1.686***	**-6.321*****	**-5.354*****
		(4.24e-01)	**(7.69e-01)**	**(9.12e-01)**	**(7.78e-01)**	**(1.015)**	**(6.60e-01)**	**(8.92e-01)**
/cut2		**9.49e-01*****	4.52e-01	1.50e-01	**1.795****		**1.086****	**2.468*****
		(2.52e-01)	(4.81e-01)	(6.45e-01)	**(7.03e-01)**		**(4.99e-01)**	**(6.65e-01)**
Observations		2,502	1,047	627	468	369	1,146	655
Log-Likelihood		-1389.9824	-488.9607	-275.2702	-232.0771	-115.0512	-442.4215	-239.0167
Prediction Error		0.5563	0.4685	0.4417	0.4983	0.3204	0.3862	0.3677

VARIABLES		Cnt=16~20	Cnt=21~30	Cnt>31	ALL
Age		-3.67e-03	**4.35e-02*** **	**3.32e-02** **	**-1.03e-02*** **
		(1.74e-02)	**(1.56e-02)**	**(1.61e-02)**	**(3.27e-03)**
Gender		**-1.622*** **	**-9.86e-01*** **	**-1.791*** **	**-5.24e-01*** **
		(3.43e-01)	**(2.89e-01)**	**(3.26e-01)**	**(6.31e-02)**
Weekend		-3.41e-01	-3.68e-01	-6.44e-01	4.89e-02
		(4.73e-01)	(4.56e-01)	(4.36e-01)	(7.50e-02)
Time of the Day	Daytime (10am-4pm)	**-8.98e-01** **	**-9.31e-01*** **	**5.90e-02**	**-1.48e-01** **
		(3.83e-01)	**(3.13e-01)**	**(3.99e-01)**	**(6.84e-02)**
	Rush Hour (7-9am, 5-8pm)	0.384	-0.0749	0.880	0.0880
		(5.12e-01)	(4.44e-01)	(6.60e-01)	(9.99e-02)
	Evening (9-11pm)	(reference)			
	Late Night (12am-3am)	**-1.266***	**-1.668** **	-1.27e-01	**-2.47e-01** **
		(6.53-01)	**(7.07e-01)**	(4.86e-01)	**(1.23e-01)**
	Early Morning (4-6am)	-5.99e-01	**-3.179*** **	-9.59e-01	-2.13e-01
		(6.80e-01)	**(1.191)**	(1.078)	(1.43-01)
WaitTime		-2.37e-04	1.06e-04	-1.18e-04	**-3.07e-04*** **
		(4.14e-04)	(4.85e-04)	(4.76e-04)	**(7.35e-05)**
Distance	Gain	-1.85e-02	7.91e-02	-1.18e-01	-2.09e-02
		(1.01e-01)	(9.50e-02)	(1.02e-01)	(1.60e-02)

	Loss	7.32e-02	5.25e-02	-1.03e-01	4.49e-02
		(9.79e-02)	(3.29e-02)	(1.45e-01)	(5.93e-03)
RideTime	Gain	-9.90e-04	3.40e-04	**2.06e-03*** **	1.25e-04
		(9.12e-04)	(7.80e-04)	**(6.63e-04)**	(1.42e-04)
	Loss	-4.18e-04	-4.22e-04	-5.67e-04	**-2.66e-04** **
		(8.76e-04)	(1.01e-03)	(1.24e-03)	**(1.08e-04)**
Cost	Gain	3.60e-04	-7.19e-05	2.06e-04	2.99e-05
		(2.72e-04)	(2.05e-04)	(1.49e-04)	(2.84e-05)
	Loss	-4.63e-05	-3.68e-05	-2.80e-04	2.65e-06
		(1.90e-04)	(2.10e-04)	(2.98e-04)	(1.19e-05)
/cut1		**-6.796*** **	**-5.738*** **	**-7.494*** **	**-6.185*** **
		(1.039)	**(1.029)**	**(1.455)**	**(2.26e-01)**
/cut2		5.71e-01	**2.658*** **	**2.840*** **	**0.959*** **
		(8.16e-01)	**(7.40e-01)**	**(7.84e-01)**	**(1.54e-01)**
Observations		417	555	778	8,564
Log-Likelihood		-151.4939	-187.6789	-197.4161	-3913.6313
Prediction Error		0.3777	0.3588	0.2576	0.4567

*** p<0.01, ** p<0.05, * p<0.1; Standard errors are in parenthesis

Table 16. Summary of significant variables for satisfaction with repeated use (DE)

VARIABLES		Cnt=1	Cnt=2	Cnt=3	Cnt=4	Cnt=5	Cnt= 6~10	Cnt= 11~15	Cnt= 16~20	Cnt= 21~30	Cnt>31	ALL
Age			(-)					(+)		(+)	(+)	(-)
Gender		(-)	(-)	(-)		(-)	(-)	(-)	(-)	(-)	(-)	(-)
Weekend												
Time of the Day	Daytime								(-)	(-)	(+)	(-)
	Rush Hour											
	Evening						(reference)					
	Late Night								(-)	(-)		(-)
	Early Morning									(-)		
WaitTime		(-)	(-)		(-)	(-)	(-)					(-)
Distance	Gain		(-)									
	Loss						(+)	(+)				
RideTime	Gain	(+)		(-)							(+)	
	Loss						(-)	(-)				(-)
Cost	Gain											
	Loss	(-)	(+)					(-)				

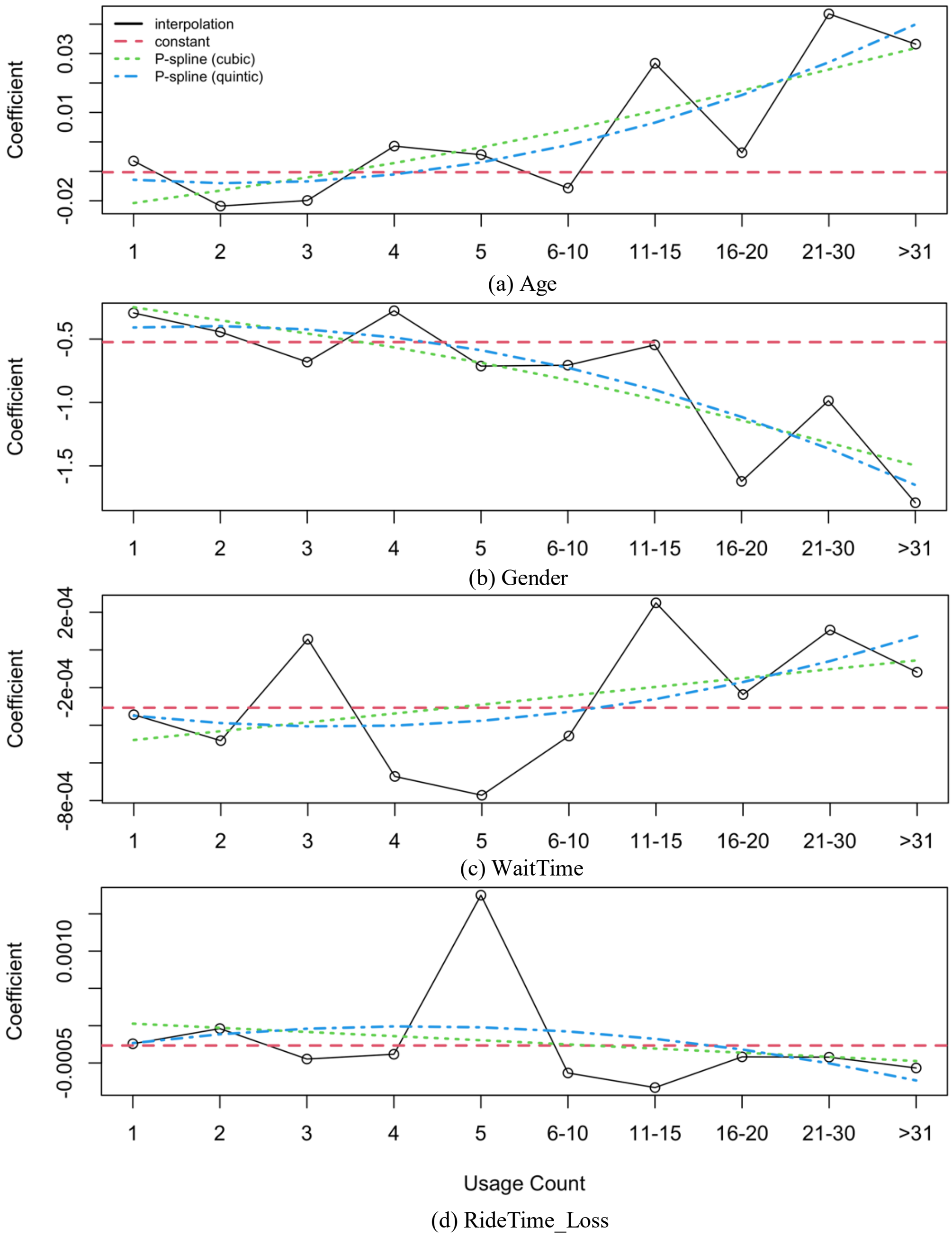
The results show that overall, the coefficients for the covariates Age, $Gender$, $Daytime$, $LateNight$, $WaitTime$, and $RideTime\_Loss$ have a significant influence on user satisfaction with the service. The Age, $Gender$, and $WaitTime$ were also almost always consistently significant. It is worth noting that the control variables that are used to determine the seasonal effect, such as $Weekend$ and $Time\_Day$, were almost always not significant in this model. The coefficients for the covariates that were consistently significant and varies by number of uses were fit using cubic and quintic splines. The result of fit is shown in Figure 27.

The graphs depict estimated coefficient with usage count τ on the horizontal axis and the estimated coefficients for the covariate effects on the vertical axis. The solid line represents the full interpolation of the estimated coefficients, and the constant coefficient estimated for the average effect is also shown. Cubic and quintic splines are those with the two lowest degrees that allow separate control of the two endpoints and the two end derivatives, while they have the lowest degree that allows reflection points. While cubic splines are the most popular because they have lowest degree, we also incorporate quintic splines to obtain a smoother curve at the expense of additional derivatives.

The figure shows that the younger the consumer, the higher the satisfaction when using the service for the first time. However, the positive effect of young age on user satisfaction decreased as the number of uses increased, and around the 5th time of use, the direction of influence completely reversed, i.e., the older the user, the higher the satisfaction with the service. In addition, satisfaction with the service was higher for men than for women, and this tendency was more heavily influenced as the number of uses increased. Lastly, a reduction in wait time and ride time increased satisfaction for all users; however, the influence of the two time covariates varied with increasing number of uses. The results suggest that the influence of shortening wait time with respect to expectation on service satisfaction decreased as usage experience accumulates for consumers, whereas the influence of shortening ride time on satisfaction with service slightly increased.

Such result indicates that towards the initial use of service, the reduction of wait time and the decrease in ride time loss with respect to expectations is what the managerial algorithms should focus on to increase service satisfaction, whereas for those who have used the service for several times and have formed a form of loyalty, matching a ride that can decrease the ride time loss with respect to expectations is of greater importance.

Following, the coverage of the 95% Bayesian confidence interval (CI) for each smoothing was calculated using Eq. (44). Figure 28 shows the estimated functional relationship as well as the simulated coverage of the 95% Bayesian CIs. To add, the summary of fit of the estimated splines is summarized in Table 17 and 18, each showing the approximate significance of parametric and nonparametric effects, respectively. Statistical inference is conducted via (approximate) frequentist chi-square tests using the Bayesian interpretation of a smoothing spline (Nychka, 1988; Wahba, 1983).



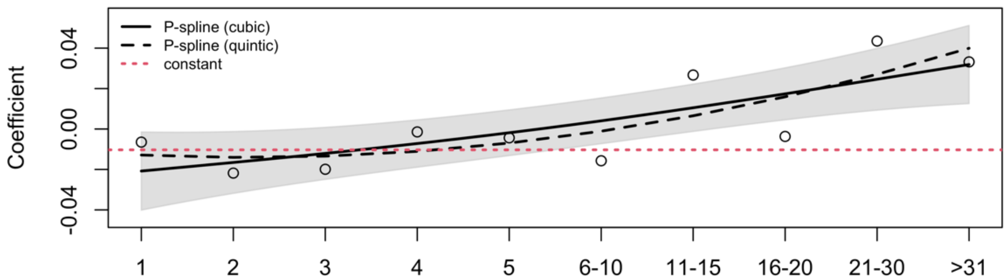
(a) Age

(b) Gender

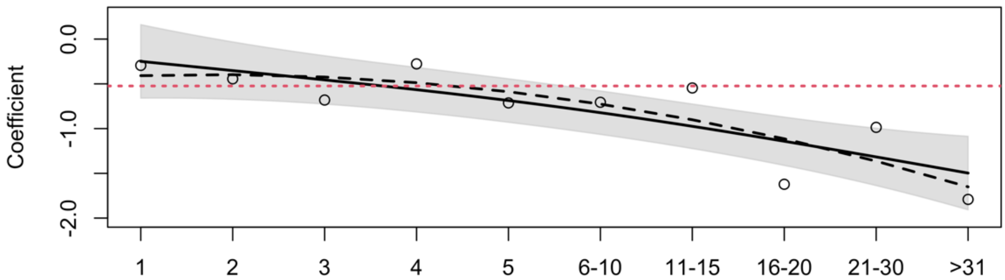
(c) WaitTime

(d) RideTime_Loss

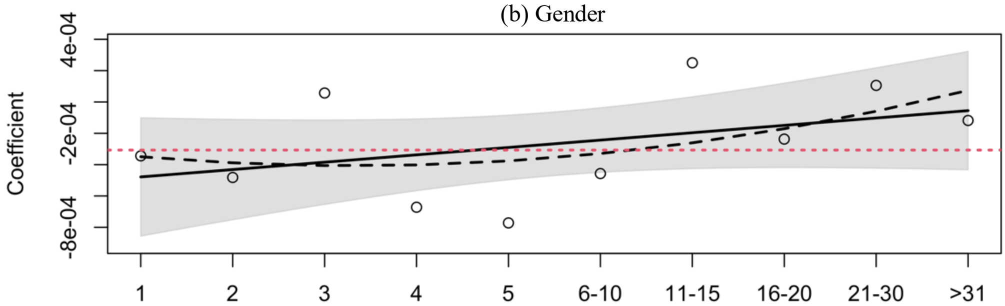
Figure 27. P-spline fits for satisfaction parameters with DE



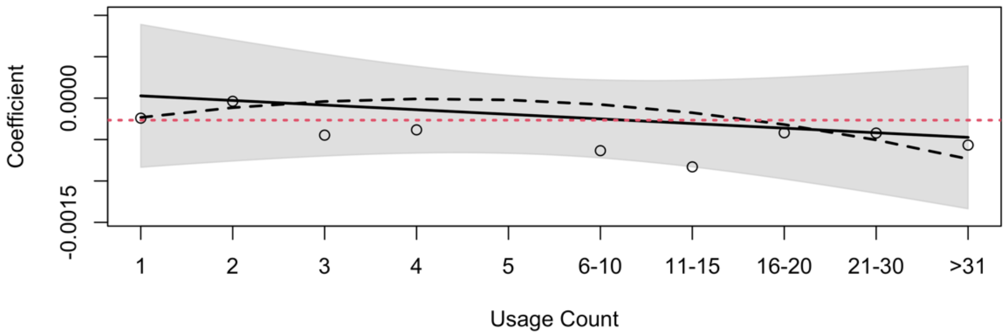
(a) Age



(b) Gender



(c) WaitTime



(d) RideTime_Loss

Figure 28. P-spline CIs for satisfaction parameters with DE

The way the output of our approach to fitting Generalized Additive Models (GAMs) is structured is to group the linear parts of the smoothers in with the other parametric terms. The reason the smooth terms are separated into two types of effect is that the results allow the researchers to decide if a smooth term has a nonlinear effect and/or a linear effect.

The parametric part refers to the linear effect of the covariate involved in the smooth. The non-parametric part refers to the nonlinearity beyond the linear/parametric part of the smooth. These tests in the nonparametric section can be interpreted as test of the null hypothesis of a linear relationship instead of a nonlinear relationship. If the nonparametric part is significant, it suggests that a linear effect of that covariate is not supported by the data. If insignificant, the linear effect needs to be considered. Presenting the information this way allows researchers to see what might be linear or effectively linear effects of covariates that we represented via smooth functions in setting up the model.

Putting the results from Figure 28 together with Tables 17 and 18, it is noticeable that for the two covariates Age and $Gender$, the coefficient estimated for the average effect is not entirely within the Bayesian CI of the cubic spline. This suggests that the nonparametric model should be preferred when fitting the given coefficients that vary with time. However, since all the nonparametric effects of the covariates are insignificant, this indicates that the linear effects should be considered. This is because the time variances are caused by long-term fluctuations.

Contrastingly, we obtain rather smooth curves for covariates $WaitTime$ and $RideTime\_Loss$, again suggesting long-term trends in their time-varying influence on satisfaction. Coefficients estimated for the average effect completely falls within the Bayesian CIs of the cubic spline, indicating that all models sufficiently explain the data, with the nonparametric model showing the long-term variance of time. Likewise, because all nonparametric effects of the covariates are insignificant, this indicates that that the linear effects need to be considered. In all, there are no short-term trends in the effect of covariates on the usage interval of ride-hailing platform's customers, and their preferences are rather consistent with time.

Table 17. Parametric effects from P-spline fits of satisfaction parameters with DE

Variable	Parameter	Cubic P-Spline				Quintic P-Spline			
		Estimate	Std. Error	t-value	Pr(>$\|t\|$)	Estimate	Std. Error	t-value	Pr(>$\|t\|$)
Age	(Intercept)	2.69e-03	4.85e-03	5.56e-01	5.94e-01	1.69e-03	4.83e-03	3.50e-01	7.37e-01
	x	5.27e-02	1.53e-02	3.43e+00	9.47e-03**	5.28e-02	1.48e-02	3.57e+00	9.12e-03**
	x^2	-	-	-	-	1.42e-01	1.05e-01	1.35e+00	2.19e-01
Gender	(Intercept)	-7.98e-01	1.01e-01	-7.89e+00	6.37e-05***	-7.78e-01	1.01e-01	-7.68e+00	1.18e-04***
	x	-1.25e+00	3.21e-01	-3.89e+00	5.11e-03**	-1.24e+00	3.11e-01	-4.00e+00	5.21e-03**
	x^2	-	-	-	-	-3.02e+00	2.21e+00	-1.37e+00	2.14e-01
WaitTime	(Intercept)	-2.67e-04	1.04e-04	-2.58e+00	3.29e-02*	-2.83e-04	1.09e-04	-2.60e+00	3.57e-02*
	x	4.23e-04	3.25e-04	1.30e+00	2.29e-01	4.23e-04	3.35e-04	1.26e+00	2.46e-01
	x^2	-	-	-	-	1.74e-03	2.38e-03	7.31e-01	4.88e-01

RideTime _Loss	(Intercept)	-2.23e-04	2.37e-04	-9.43e-01	3.73e-01	-1.91e-04	2.51e-04	-7.59e-01	4.73e-01
	x	-5.01e-04	7.42e-04	-6.76e-01	5.18e-01	-5.01e-04	7.70e-04	-6.51e-01	5.36e-01
	x^2	-	-	-	-	-3.52e-03	5.48e-03	-6.52e-01	5.41e-01

*** p<0.001, ** p<0.01, * p<0.05, . p<0.1

Table 18. Nonparam. effects from P-spline fits of satisfaction parameters with DE

		Cubic P-Spline					Quintic P-Spline				
Variable	Parameter	DF	Sums of Squares	Mean Squares	F-value	Pr(>F)	DF	Sums of Squares	Mean Squares	F-value	Pr(>F)
Age	$f(x)$	3.18e-01	1.04e-04	3.27e-04	1.40e+00	2.72e-01	6.62e-09	1.79e-13	2.70e-05	1.21e-01	7.38e-01
	Residuals	7.18e+00	1.79e-03	2.23e-04	-	-	7.00e+00	1.56e-03	2.23e-04	-	-
Gender	$f(x)$	3.94e-01	6.10e-02	1.55e-01	1.53e+00	2.52e-01	6.62e-09	1.33e-10	2.02e-02	2.05e-01	6.65e-01
	Residuals	7.61e+00	7.69e-01	1.01e-01	-	-	7.00e+00	6.88e-01	9.84e-02	-	-

WaitTime	$f(x)$	7.68e-07	6.11e-14	7.95e-08	7.40e091	4.15e-01	6.62e-09	4.33e-16	6.54e-08	5.73e-01	4.74e-01
	Residuals	8.00e+00	8.60e-07	1.07e-07	-	-	7.00e+00	7.98e-07	1.14e-07	-	-
RideTime _Loss	$f(x)$	7.68e-07	2.29e-13	2.97e-07	5.31e-01	4.87e-01	6.62e-09	1.19e-15	1.79e-07	2.97e-01	6.03e-01
	Residuals	8.00e+00	4.48e-06	5.60e-07	-	-	7.00e+00	4.23e-06	6.05e-07	-	-

*** p<0.001, ** p<0.01, * p<0.05, . p<0.1

5.4.2.2 Expectations Updating Effect

This section analyzes the time-varying effects of covariates in explaining consumers' use of ride-hailing platforms based on updated expectations, assuming that consumers' service expectations are updated with repeated use. As in the previous analysis based on described expectations, Macaron Taxi data, which included a total of 8,564 successful immediate rides, were categorized by the number of experiences with the service.

To determine the overall impact of the updated expectations, the expected errors from the (described) service expectations are first calculated from the person's usage history, i.e., their usage experience, as shown in Figure 29. In doing so, it is reasonable to give more weights to recent observations than to observations from the distant past. The forecasts are therefore calculated using weighted averages, with the weights decreasing exponentially as the observations become more distant – the smallest weights are associated with the oldest observations. Such exponential smoothing involves the smoothing coefficient θ_{ref} with a value between 0 and 1. A value close to 0 indicates that observations from further in the past are weighted more heavily, and a value close to 1 indicates that more weight is given to the recent observations.

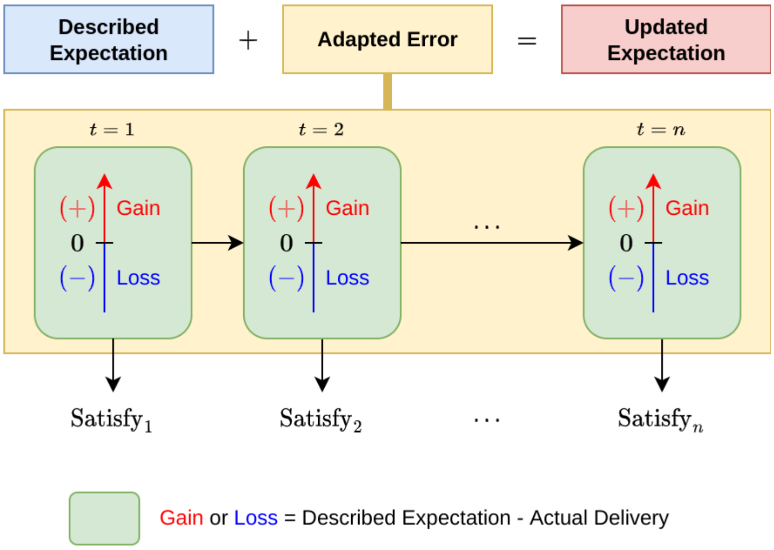


Figure 29. Formation of described and updated expectations

In this section, the exponential smoothing parameter of 0.5 is used for all three components: $\theta_{ref,dist}, \theta_{ref,time}, \theta_{ref,cost}$. The updated service expectations are as calculated using Eq. (78) to Eq. (80).

$$UpdatedDistE_{ijt} = DistSysE_{ijt} + \{\theta_{ref,dist} \cdot (DistGain_{ij,t-1} - DistLoss_{ij,t-1}) + (1 - \theta_{ref,dist}) \cdot ExptErrorDist_{ij,t-1}\} \cdots\cdots\cdots\cdots \text{Eq. (78)}$$

$$UpdatedTimeE_{ijt} = DisTimeE_{ijt} + \{\theta_{ref,time} \cdot (TimeGain_{ij,t-1} - TimeLoss_{ij,t-1}) + (1 - \theta_{ref,time}) \cdot ExptErrorTime_{ij,t-1}\} \cdots\cdots\cdots\cdots \text{Eq. (79)}$$

$$UpdatedCostE_{ijt} = CostSysE_{ijt} + \{\theta_{ref,cost} \cdot (CostGain_{ij,t-1} - CostLoss_{ij,t-1}) + (1 - \theta_{ref,cost}) \cdot ExptErrorCost_{ij,t-1}\} \cdots\cdots\cdots\cdots \text{Eq. (80)}$$

The empirical model used to analyze the accumulated experience effects of covariates on service satisfaction ($Satisfy$) of consumers in the Macaron ride-hailing platform service is presented in Eq. (81). The estimation results using updated expectations with exponential smoothing parameter values of 0.5 are presented in Table 19 below, and the significant variables for each use are again summarized in Table 20.

$$
\begin{aligned}
Satisfy_{i,t} = {} & \beta_0 + \beta_1 Age_i + \beta_2 Gender_i + \beta_3 WaitTime_{i,t} \\
& + I(UpdatedDistE_{i,t} \geq RealDist_{i,t})\beta_4(UpdatedDistE_{i,t} - RealDist_{i,t}) \\
& + I(UpdatedDistE_{i,t} < RealDist_{i,t})\beta_5|UpdatedDistE_{i,t} - RealDist_{i,t}| \\
& + I(UpdatedTimeE_{i,t} \geq RealTime_{i,t})\beta_6(UpdatedTimeE_{i,t} - RealRide_{i,t}) \\
& + I(UpdatedTimeE_{i,t} < RealRide_{n,t})\beta_7|UpdatedTimeE_{i,t} - RealRide_{i,t}| \\
& + I(UpdatedCostE_{i,t} \geq RealCost_{i,t})\beta_8(UpdatedCostE_{i,t} - RealCost_{i,t}) \\
& + I(UpdatedCostE_{i,t} < RealCost_{i,t})\beta_7|UpdatedCostE_{i,t} - RealCost_{,t}| \\
& + \epsilon_{it} \qquad \cdots\cdots\cdots \text{Eq. (81)}
\end{aligned}
$$

Coefficients for covariates as identical to those of described expectations analysis were fit using both the cubic and quintic P-spline functions as shown in Figure 30. From the figure, it can be seen that the coefficients for all four covariates show similar long-time trends, only to note that the positive the effect of reduced ride time loss with respect to expectations on service satisfaction increased more steeply with increasing number of uses.

The coverage of the 95% Bayesian confidence interval (CI) for each smoothing was calculated using Eq. (44). Figure 31 shows the estimated functional relationship as well as the simulated coverage of the 95% Bayesian CIs. The summary of fit of the estimated splines is summarized in Table 21 and 22.

Table 19. Estimation results for satisfaction with UE (smoothing parameter=0.5)

VARIABLES		Cnt=1	Cnt=2	Cnt=3	Cnt=4	Cnt=5	Cnt=6~10	Cnt=11~15
Age		-8.00e-03	**-2.17e-02****	-1.65e-02	-1.83e-04	-7.26e-03	-1.66e-02	**2.64e-02***
		(5.38e-03)	**(1.01e-02)**	(1.34e-02)	(1.41e-02)	(2.01e-02)	(1.04e-02)	**(1.37e-02)**
Gender		**-3.19e-01*****	**-4.07e-01****	**-7.32e-01*****	-3.78e-01	**-6.64e-01***	**-7.94e-01*****	**-5.83e-01****
		(1.02e-01)	**(1.80e-01)**	**(2.43e-01)**	(2.69e-01)	**(3.64e-01)**	**(1.91e-01)**	**(2.56e-01)**
WaitTime		**-3.62e-04*****	**-4.94e-04****	**1.29e-04**	**-5.70e-04***	**-7.59e-04***	**-4.84e-04***	2.56e-04
		(1.21e-04)	**(2.01e-04)**	**(2.40e-04)**	**(3.47e-04)**	**(4.50e-04)**	**(2.49e-04)**	(4.16e-04)
Distance	Gain	-1.75e-02	**-1.01e-01****	-9.44e-02	-2.38e-02	-1.50e-01	-2.37e-02	1.44e-02
		(1.92e-02)	**(3.95e-02)**	(6.81e-02)	(6.76e-02)	(1.23e-01)	(5.90e-02)	(6.76e-02)
	Loss	-1.68e-04	1.16e-02	-4.18e-02	-7.41e-02	-5.09e-02	**4.88e-02****	**1.71e-01*****
		(8.83e-03)	(2.00e-02)	(3.63e-02)	(4.86e-02)	(7.60e-02)	**(1.94e-02)**	**(5.27e-02)**
RideTime	Gain	**4.33e-04****	-1.45e-06	**-1.02e-03****	-2.15e-06	-3.56e-04	-5.23e-04	**-1.67e-03*****
		(2.09e-04)	(2.89e-04)	**(4.59e-04)**	(4.56e-04)	(7.48e-04)	(4.30e-04)	**(5.82e-04)**
	Loss	-2.45e-04	-2.00e-04	**-6.00e-04***	-8.91e-05	-4.56e-05	**-5.91e-04****	**-8.08e-04****
		(1.59e-04)	(1.87e-04)	**(3.20e-04)**	(3.95e-04)	(7.49e-04)	**(2.98e-04)**	**(3.95e-04)**
Cost	Gain	1.78e-05	6.40e-05	**1.99e-04****	8.31e-05	2.47e-04	**-2.09e-04***	1.61e-04
		(3.54e-05)	(6.58e-05)	**(9.54e-05)**	(1.07e-04)	(1.71e-04)	**(1.25e-04)**	(1.48e-04)
	Loss	**-4.53e-05****	**3.68e-05****	4.00e-05	1.34e-05	1.33e-04	-1.34e-07	**-1.48e-04****
		(1.92e-05)	**(1.83e-05)**	(2.47e-05)	(6.43e-05)	(1.61e-04)	(5.73e-05)	**(7.16e-05)**
/cut1		-6.125***	-6.975***	-7.049***	-4.233***	1.726*	-6.712***	-5.620***

	(4.16e-01)	(7.47e-01)	(8.83e-01)	(7.33e-01)	(9.21e-01)	(6.51e-01)	(8.88e-01)
/cut2	8.04e-01***	6.14e-01	2.98e-01	1.656**		7.89e-01*	2.352***
	(2.36e-01)	(4.45e-01)	(5.99e-01)	(6.45e-01)		(4.70e-01)	(6.27e-01)
Observations	2,502	1,047	627	468	369	1,146	655
Log-Likelihood	-1391.8091	-489.7736	-273.5431	-232.9184	-119.4861	-436.2779	-236.2949
Prediction Error	0.5563	0.4685	0.4417	0.4983	0.3204	0.3862	0.3677

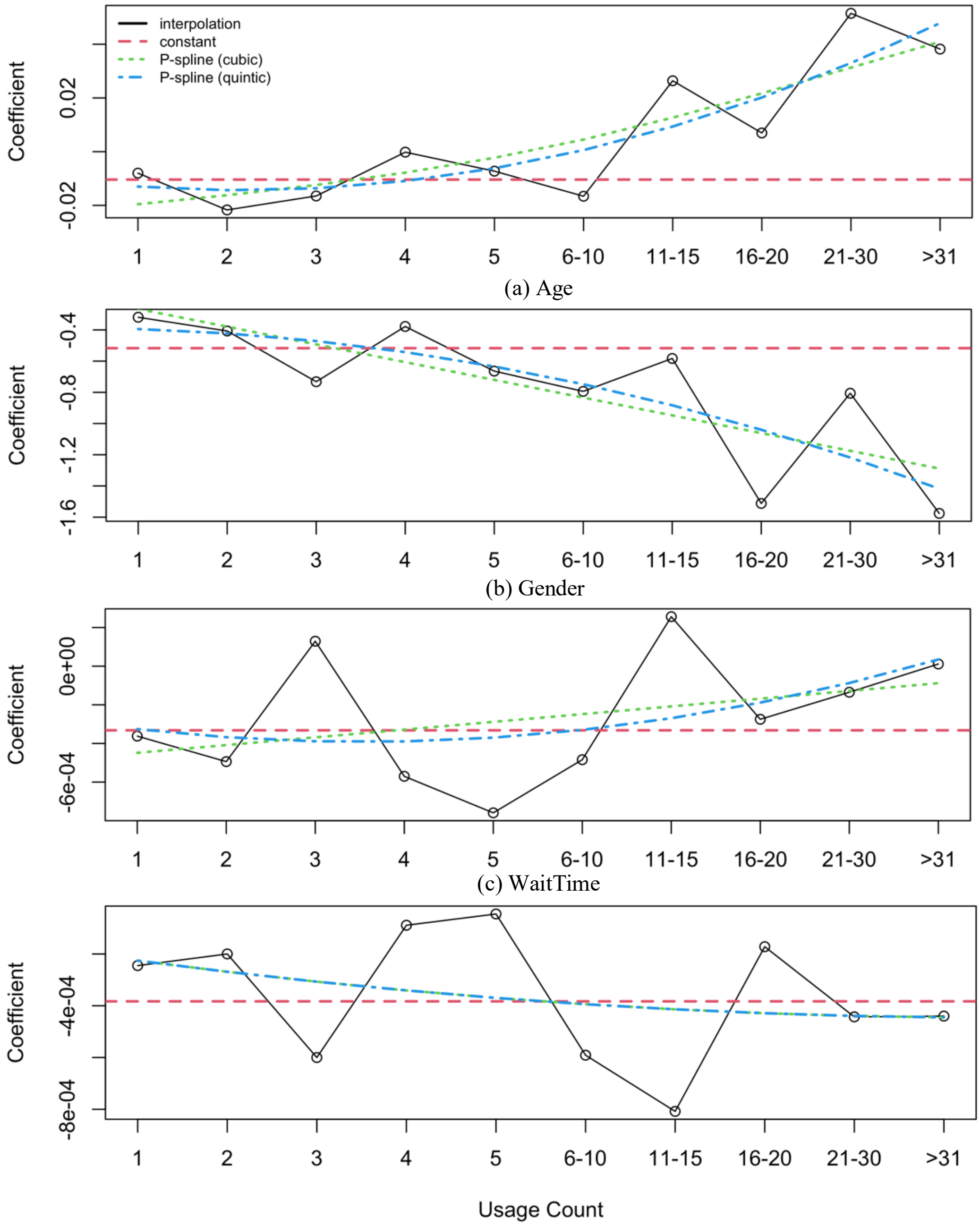
VARIABLES		Cnt=16~20	Cnt=21~30	Cnt>31	ALL
Age		6.96e-03	**5.15e-02*** **	**3.82e-02** **	**-1.04e-02*** **
		(1.72e-02)	**(1.54e-02)**	**(1.63e-02)**	**(3.24e-03)**
Gender		**-1.512*** **	**-8.06e-01*** **	**-1.576*** **	**-5.17e-01*** **
		(3.33e-01)	**(2.82e-01)**	**(3.23e-01)**	**(6.18e-02)**
WaitTime		-2.75e-04	-1.35e-04	1.11e-05	**-3.32e-04*** **
		(3.66e-04)	(4.61e-04)	(4.69e-04)	**(7.34e-05)**
Distance	Gain	-2.26e-02	4.62e-02	-3.01e-02	**-2.92e-02** **
		(7.91e-02)	(7.87e-02)	(6.99e-02)	**(1.43e-02)**
	Loss	**1.17e-01***	**8.10e-02** **	-1.74e-01	8.87e-03
		(6.80e-02)	**(3.23e-02)**	(1.27e-01)	(6.23e-03)
RideTime	Gain	-7.67e-04	5.49e-04	**1.42e-03*** **	-8.82e-05
		(6.31e-04)	(5.48e-04)	**(5.00e-04)**	(1.18e-04)
	Loss	-1.72e-04	-4.43e-04	-4.40e-04	**-3.84e-04*** **

		(5.98e-04)	(8.90e-04)	(8.90e-04)	**(9.32e-05)**
Cost	Gain	2.06e-04	**-3.51e-04***	6.52e-05	2.87e-05
		(1.79e-04)	**(1.80e-04)**	(1.06e-04)	(2.55e-05)
	Loss	-1.27e-04	**-3.68e-04****	-2.16e-04	-1.29e-06
		(1.20e-04)	**(1.78e-04)**	(2.13e-04)	(1.11e-05)
/cut1		-5.687***	-4.694***	-7.242***	-6.192***
		(9.09e-01)	(9.38e-01)	(1.464)	(2.19e-01)
/cut2		1.501**	3.544***	3.142***	9.64e-01***
		(7.45e-01)	(6.88e-01)	(6.91e-01)	(1.44e-01)
Observations		417	555	778	8,564
Log-Likelihood		-155.1528	-192.2037	-199.5080	-3911.1860
Prediction Error		0.3777	0.3588	0.2576	0.4567

*** p<0.01, ** p<0.05, * p<0.1; Standard errors are in parenthesis

Table 20. Summary of significant variables for satisfaction with UE (smoothing parameter=0.5)

VARIABLES		Cnt=1	Cnt=2	Cnt=3	Cnt=4	Cnt=5	Cnt= 6~10	Cnt= 11~15	Cnt= 16~20	Cnt= 21~30	Cnt>31	ALL
Age			(-)					(+)		(+)	(+)	(-)
Gender		(-)	(-)	(-)		(-)	(-)	(-)	(-)	(-)	(-)	(-)
WaitTime		(-)	(-)	(+)	(-)	(-)	(-)					(-)
Distance	Gain		(-)									(-)
	Loss						(+)	(+)	(+)	(+)		
RideTime	Gain	(+)		(-)				(-)			(+)	
	Loss			(-)			(-)	(-)				(-)
Cost	Gain			(+)			(-)			(-)		
	Loss	(-)	(+)					(-)		(-)		



(a) Age

(b) Gender

(c) WaitTime

(d) RideTime_Loss

Figure 30. P-spline fits for satisfaction parameters with UE (smoothing param=0.5)

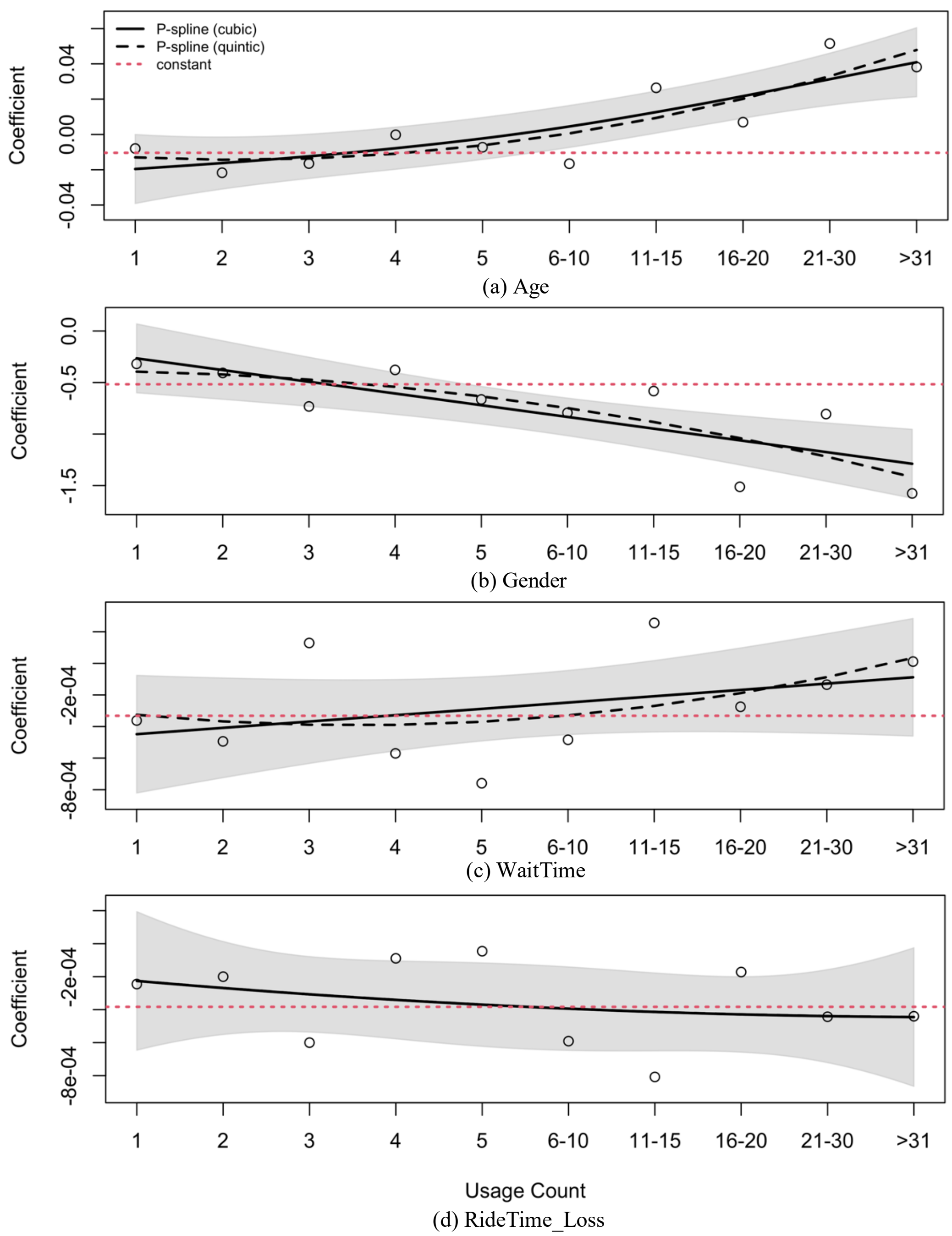


Figure 31. P-spline CIs for satisfaction parameters with UE (smoothing param=0.5)

Table 21. Parametric effects from P-spline fits of satisfaction parameters with UE (smoothing parameter=0.5)

		Cubic P-Spline				Quintic P-Spline			
Variable	Parameter	Estimate	Std. Error	t-value	Pr(>|t|)	Estimate	Std. Error	t-value	Pr(>|t|)
Age	(Intercept)	4.62e-03	4.59e-03	1.01e+00	3.46e-01	3.87e-03	4.57e-03	8.24e-01	4.37e-01
	x	6.05e-02	1.48e-02	4.10e+00	4.08e-03**	6.09e-02	1.40e-02	4.34e+00	3.38e-03**
	x^2	-	-	-	-	1.64e-01	9.98e-02	1.65e+00	1.44e-01
Gender	(Intercept)	-7.77e-01	9.17e-02	-8.48e+00	2.88e-05***	-7.51e-01	9.55e-02	-7.97e+00	9.31e-05***
	x	-1.02e+00	2.87e-01	-3.56e+00	7.40e-03***	-1.02e+00	2.93e-01	-3.49e+00	1.01e-02*
	x^2	-	-	-	-	-1.74e+00	2.08e+00	-8.35e-01	4.31e-01
WaitTime	(Intercept)	-2.68e-04	1.02e-04	-2.63e+00	3.03e-02*	-2.84e-04	1.08e-04	-2.63e+00	3.38e-02*
	x	3.60e-04	3.20e-04	1.13e+00	2.93e-01	3.60e-04	3.31e-04	1.09e+00	3.12e-01
	x^2	-	-	-	-	1.65e-03	2.35e-03	7.03e-01	5.05e-01
	(Intercept)	-3.65e-04	2.36e-05	15.47e+00	4.32e-01	-3.67e-04	8.80e-05	-4.17e+00	4.19e-03**

RideTime_Loss	x	-1.99e-04	7.24e-05	-2.74e+00	4.32e-01	-2.19e-04	2.70e-04	-8.12e-01	4.44e-01
	x^2	-	-	-	-	3.77e-04	1.92e-03	1.96e-01	8.50e-01

*** p<0.001, ** p<0.01, * p<0.05, . p<0.1

Table 22. Nonparam. effects from P-spline fits of satisfaction parameters with UE (smoothing parameter=0.5)

		Cubic P-Spline					Quintic P-Spline				
Variable	Parameter	DF	Sums of Squares	Mean Squares	F-value	Pr(>F)	DF	Sums of Squares	Mean Squares	F-value	Pr(>F)
Age	$f(x)$	6.11e-01	2.65e-04	4.33e-04	2.09e+00	1.90e-01	6.62e-09	2.36e-13	3.58e-05	1.79e-01	6.85e-01
	Residuals	7.399	1.53e-03	2.07e-04	-	-	7.00e+00	1.40e-03	2.00e-04	-	-
Gender	$f(x)$	7.68e-07	4.22e-08	5.50e-02	6.54e-01	4.42e-01	6.62e-09	9.11e-11	1.38e-02	1.58e-01	7.03e-01
	Residuals	8.00e+00	6.73e-01	8.41e-02	-	-	7.00e+00	6.12e-01	8.74e-02	-	-
WaitTime	$f(x)$	7.68e-07	4.82e-14	6.28e-08	6.02e-01	4.60e-01	6.62e-09	8.89e-17	1.34e-08	1.21e-01	7.39e-01
	Residuals	8.00e+00	8.35e-07	1.04e-07	-	-	7.00e+00	7.79e-07	1.11e-07	-	-

RideTime _Loss	$f(x)$	7.80e+00	5.10e-07	6.54e-08	24.88e+00	5.96e-01	6.62e-09	9.91e-17	1.50e-08	2.02e-01	6.68e-01
	Residuals	2.02e-01	5.29e-10	2.63e-09	-	-	7.00e+00	5.20e-07	7.42e-08	-	-

*** p<0.001, ** p<0.01, * p<0.05, . p<0.1

5.4.2.2.1 Expectation Updating Parameter Estimation

In the previous section, the exponential smoothing parameter of 0.5 was used for all three components of the expectation update: $\theta_{ref,dist}, \theta_{ref,time}, \theta_{ref,cost}$, while the problem with this method is to determine the optimal parameters to minimize the prediction error. In this section, the Limited-memory Broyden–Fletcher–Goldfarb–Shanno optimizer (L-BFGS) using pytorch is used to determine the smoothing coefficients $\theta_{ref,dist}, \theta_{ref,time}, \theta_{ref,cost}$ that minimize prediction error. Pytorch-LBFGS is a modular implementation of L-BFGS, a popular quasi-Newton method. The optimized smoothing parameter estimates are shown in Table 23.

While optimizers are algorithms or methods used to change the attributes of the neural network such as weights and learning rate in order to reduce the losses, loss function in a neural network quantifies the difference between the expected outcome and the outcome produced by the deep learning model. In other words, it is used to quantify how good or bad the model is performing. Therefore, it is notable that the greater the loss (train error), the greater the error in the prediction.

Table 23. Estimation results for service satisfaction with UE (optimized smoothing parameter)

VARIABLES		Cnt=1	Cnt=2	Cnt=3	Cnt=4	Cnt=5	Cnt=6~10	Cnt=11~15
Age		-8.00e-03	-2.17e-02	-1.69e-02	5.88e-04	-1.04e-02	-1.66e-02	2.68e-02
Gender		-3.19e-01	-4.07e-01	-7.33e-01	-3.82e-01	-6.33e-01	-8.02e-01	-5.80e-01
WaitTime		-3.62e-04	-4.93e-04	1.37e-04	-5.57e-04	-7.65e-04	-4.89e-04	2.56e-04
Distance	Gain	-1.75e-02	-1.01e-01	-1.41e-01	-1.42e-02	-3.10e-01	-3.32e-02	-3.98e-03
	Loss	-1.70e-04	1.16e-02	-4.20e-02	-8.92e-02	-6.06e-03	4.23e-02	1.80e-01
RideTime	Gain	4.33e-04	-1.41e-06	-1.08e-03	3.53e-05	-5.11e-04	-5.02e-04	-1.77e-03
	Loss	-2.45e-04	-2.00e-04	-6.23e-04	-5.82e-05	-7.51e-05	-5.76e-04	-8.65e-04
Cost	Gain	1.78e-05	6.40e-05	2.34e-04	8.74e-05	2.68e-04	-2.04e-04	2.35e-04
	Loss	-4.53e-05	3.68e-05	4.24e-05	1.75e-05	1.37e-04	-8.36e-06	-1.11e-04
$\theta_{ref,dist}$		5.00e-01	5.00e-01	9.85e-02	7.71e-01	1.00e+00	2.72e-01	6.30e-01
$\theta_{ref,time}$		5.00e-01	5.00e-01	3.443-01	7.83e-01	6.87e-01	5.42e-01	4.38e-01
$\theta_{ref,cost}$		5.00e-01	5.00e-01	3.72e-01	6.65e-01	7.23e-01	3.74e-01	8.90e-02
/cut1		-5.7059	-5.9737	-5.5925	-4.1369	-64.5065	-5.4045	-5.7779
/cut2		1.2229	1.6144	1.8237	1.7650	2.3759	2.1018	2.2027
Observations		2,502	1,047	627	468	369	1,146	655
Train Error		0.5563	0.4678	0.4332	0.4969	0.3187	0.3805	0.3597

VARIABLES		Cnt=16~20	Cnt=21~30	Cnt>31	ALL
Age		4.59e-03	5.23e-02	3.29e-02	-1.04e-02
Gender		-1.51e+00	-9.59e-01	-1.47e+00	-5.16e-01
WaitTime		-2.35e-04	3.30e-05	-2.28e-05	-3.34e-04
Distance	Gain	-7.71e-02	8.36e-02	-5.65e-03	-2.41e-02
	Loss	1.32e-01	8.60e-02	-1.52e-01	9.78e-03
RideTime	Gain	-1.31e-03	-1.10e-03	1.61e-03	-1.33e-04
	Loss	-2.92e-04	-5.99e-04	-6.07e-04	-3.41e-04
Cost	Gain	3.67e-04	-2.07e-04	-5.83e-05	3.37e-05
	Loss	-1.44e-04	-3.99e-04	-3.44e-04	-2.23e-05
$\theta_{ref,dist}$		7.48e-01	5.00e-01	0	1.72e-05
$\theta_{ref,time}$		8.80e-01	1.00e+00	8.34e-01	4.72e-16
$\theta_{ref,cost}$		9.80e-01	3.99e-07	1.00e+00	9.99e-01
/cut1		-4.9632	-6.4258	-7.8455	-5.4604
/cut2		2.3970	2.2043	2.7703	1.7047
Observations		417	555	778	8,564
Train Error		0.3658	0.3397	0.2517	0.4563

5.4.2.2.2 Updated Preference with Accumulated Usage Experience

The empirical model used to analyze the accumulated experience effects of covariates on service satisfaction ($Satisfy$) of consumers in the Macaron ride-hailing platform service is identical to that used in the previous section. The coefficient estimates using updated expectations using optimized parameters are also summarized in Table 23.

Coefficients for covariates as identical to those of previous analysis were fit using both the cubic and quintic P-spline functions as shown in Figure 32. From the figure, it can be seen that the coefficients for all four covariates show similar long-time trends, only to note that the positive the effect of reduced ride time loss with respect to expectations on service satisfaction increased more steeply with increasing number of uses compared to the model using described expectations.

The coverage of the 95% Bayesian confidence interval (CI) for each smoothing was calculated using Eq. (44). Figure 33 shows the estimated functional relationship as well as the simulated coverage of the 95% Bayesian CIs. The summary of fit of the estimated splines is summarized in Table 24 and 25.

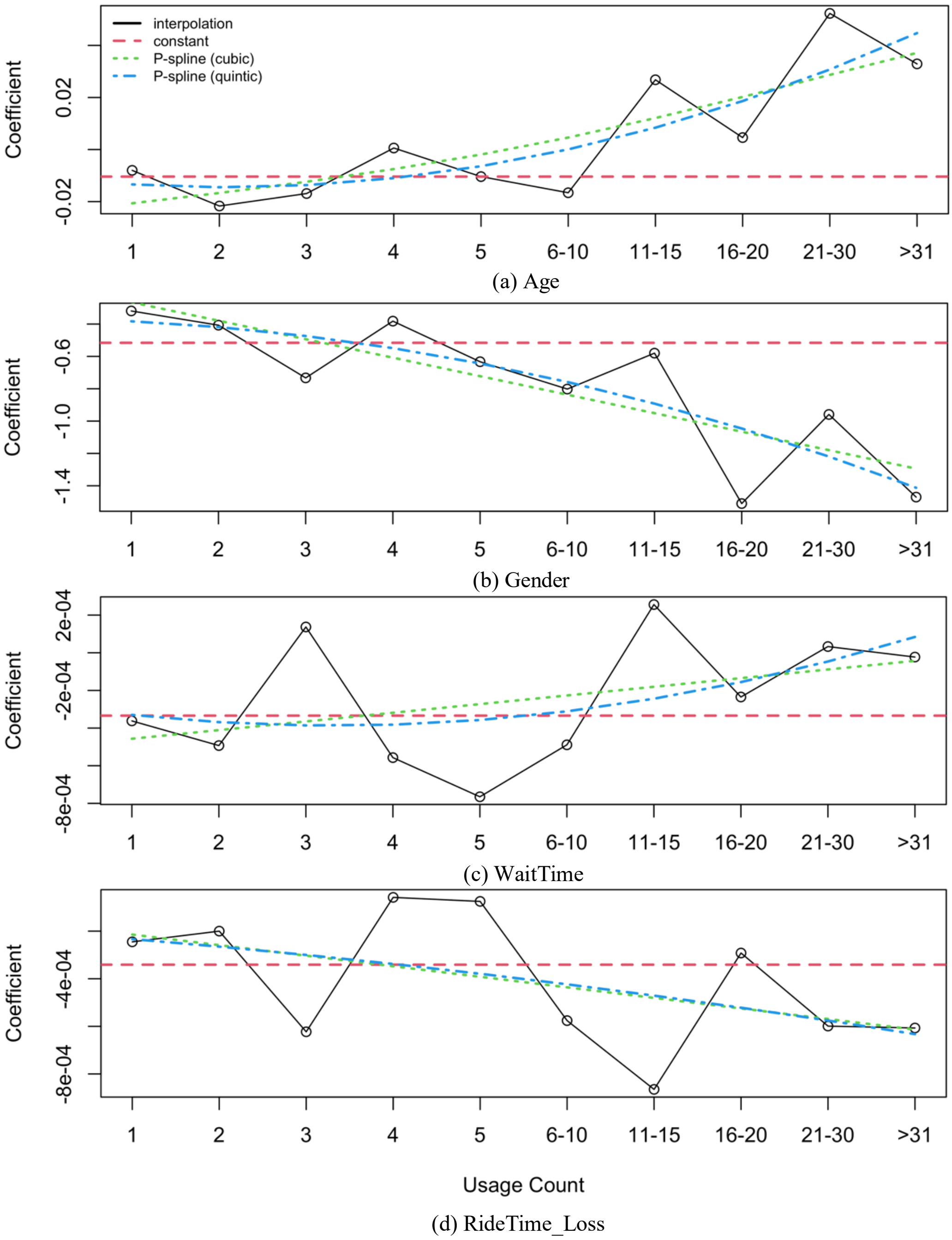
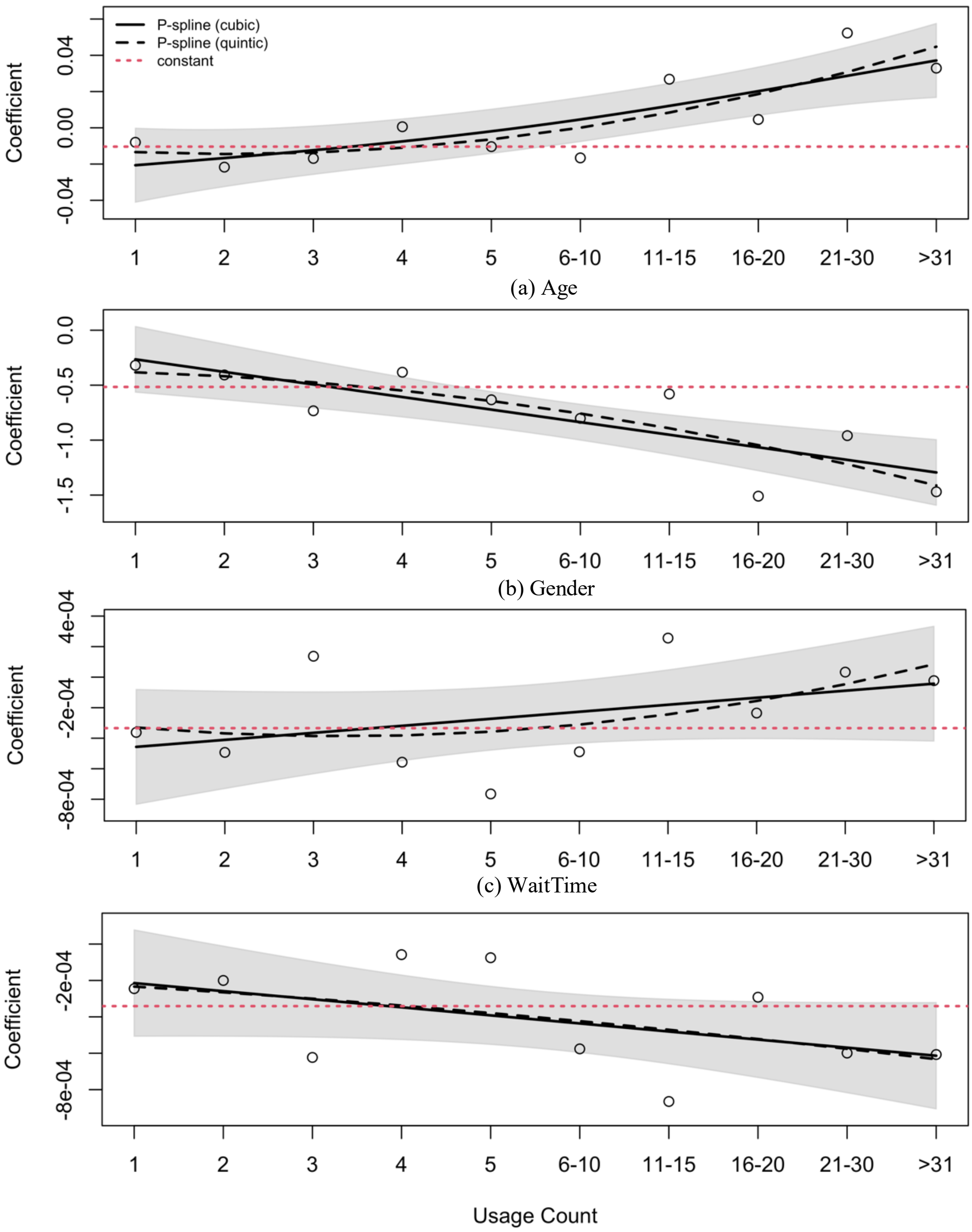


Figure 32. P-spline fits for satisfaction parameters with UE (optimized)



(a) Age

(b) Gender

(c) WaitTime

(d) RideTime_Loss

Figure 33. P-spline CIs for satisfaction parameters with UE (optimized)

Table 24. Parametric effects from P-spline fits of satisfaction parameters with UE (optimized smoothing parameter)

Variable	Parameter	Cubic P-Spline				Quintic P-Spline			
		Estimate	Std. Error	t-value	Pr(>$\|t\|$)	Estimate	Std. Error	t-value	Pr(>$\|t\|$)
Age	(Intercept)	3.88e-03	4.96e-03	7.84e-01	4.57e-01	2.94e-03	4.97e-03	5.92e-01	5.72e-01
	x	5.78e-02	1.58e-02	3.66e+00	7.06e-03**	5.81e-02	1.53e-02	3.81e+00	6.62e-03**
	x^2	-	-	-	-	1.53e-01	1.09e-01	1.41e+00	2.02e-01
Gender	(Intercept)	-7.79e-01	8.19e-02	-9.52e+00	1.22e-05***	-7.65e-01	8.50e-02	-9.00e+00	4.28e-05***
	x	-1.03e+00	2.57e-01	-4.01e+00	3.88e-03**	-1.03e+00	2.61e-01	-3.95e+00	5.56e-03**
	x^2	-	-	-	-	-1.59e+00	1.86e+00	-8.59e-01	4.19e-01
WaitTime	(Intercept)	-2.50e-04	1.03e-04	-2.42e+00	4.18e-02*	-2.66e-04	1.09e-04	-2.44e+00	4.44e-02*
	x	4.14e-04	3.23e-04	1.28e+00	2.36e-01	4.14e-04	3.33e-04	1.24e+00	2.54e-01
	x^2	-	-	-	-	1.71e-03	2.37e-03	7.21e-01	4.94e-01
RideTime	(Intercept)	-4.14e-04	8.01e-05	-5.17e+00	8.54e-04***	-4.11e-04	8.73e-05	-4.71e+00	2.17e-03**

_Loss	x	-3.99e-04	2.51e-04	-1.59e+00	1.50e-01	-3.99e-04	2.68e-04	-1.49e+00	1.80e-01
	x^2	-	-	-	-	-2.58e-04	1.91e-03	-1.36e-01	8.96e-01

*** p<0.001, ** p<0.01, * p<0.05, . p<0.1

Table 25. Nonparam. effects from P-spline fits of satisfaction parameters with UE (optimized smoothing parameter)

		Cubic P-Spline					Quintic P-Spline				
Variable	Parameter	DF	Sums of Squares	Mean Squares	F-value	Pr(>F)	DF	Sums of Squares	Mean Squares	F-value	Pr(>F)
Age	$f(x)$	4.54e-01	1.81e-04	3.99e-04	1.65e+00	2.37e-01	6.62e-09	4.15e-13	6.28e-05	2.65e-01	6.23e-01
	Residuals	7.55e+00	1.83e-03	2.42e-04	-	-	7.00e+00	1.66e-03	2.37e-04	-	-
Gender	$f(x)$	7.68e-07	3.77e-08	4.91e-02	7.32e-01	4.17e-01	6.62e-09	3.12e-11	4.71e-03	6.79e-02	8.02e-01
	Residuals	8.00e+00	5.36e-01	6.70e-02	-	-	7.00e+00	4.85e-01	6.93e-02	-	-
WaitTime	$f(x)$	7.68e-07	5.48e-14	7.14e-08	6.71e-01	4.36e-01	6.62e-09	1.59e-16	2.40e-08	2.12e-01	6.59e-01
	Residuals	8.00e+00	8.51e-07	1.06e-07	-	-	7.00e+00	7.93e-07	1.13e-07	-	-

RideTime_Loss	$f(x)$	7.68e-07	5.67e-15	7.39e-09	1.15e-01	7.43e-01	6.62e-09	1.11e-16	1.68e-08	2.30e-01	6.46e-01
	Residuals	8.00e+00	5.13e-07	6.42e-08	-	-	7.00e+00	5.12e-07	7.31e-08	-	-

*** p<0.001, ** p<0.01, * p<0.05, . p<0.1

5.4.2.3 Perception Updating Effect (Using RNN LSTM Model)

This section analyzes the time-varying effects of covariates in explaining consumers' use of ride-hailing platforms based on updated perception, assuming that individual's perception is formed relative to a reference value that has been built up through experience. In the previous section, in which consumer expectations were updated using adapted error, we observed how user satisfaction is affected by the "Service Gap", i.e., the gap between service expectation and actual service delivery. Contrastingly, in this section, consumer's updated perception of the service component is directly estimated using the LSTM model, a form of RNN, to illustrate the impact of the consumer's updated perception on service satisfaction. In other words, by directly updating the consumer's updated perception and estimating its influence on service satisfaction, we explain the effect of the "Perception Gap".

Here, the updated perception directly indicates the "consumers' perceived service gains and losses," which is similar but not identical to the updated expectation of "the difference between the updated expectation and the actual service delivery (gain/loss)." Originally, the expected gain and losses were defined as functions of previous gains and losses (i.e., $f(gain_t, loss_t, gain_{t-1}, loss_{t-1}, \dots)$), where the smoothing parameter could be optimized ("learned"). In this section, this function is more generalized in the form of LSTM. Conceptual framework for updated perception is represented in Figure 34.

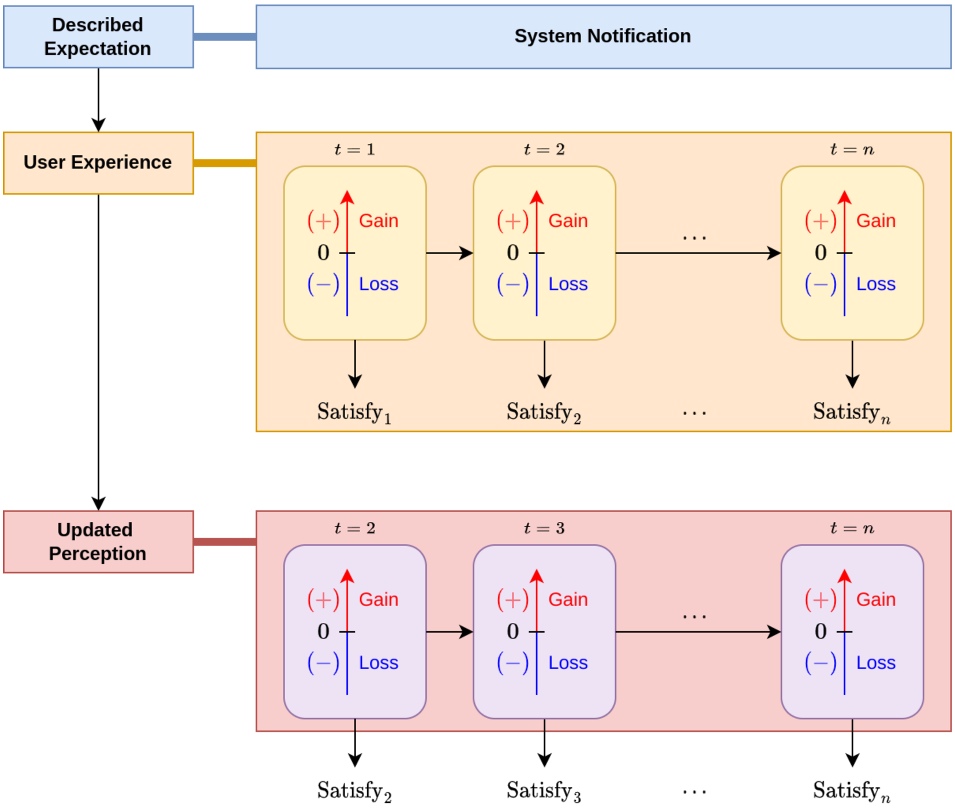


Figure 34. Formation of updated perception

In this paper, the final objective of LSTM was learning to predict the dependent variable, service satisfaction, well. Since the result of the function may be too arbitrary when using the LSTM, the output range was restricted to [-1,1] to follow the original constraint that either the gain or loss always has a value of 0, so that the output of LSTM does not differ significantly from existing gain-loss. It was also designed so that LSTM can only adjust the existing gain or loss between multiple of 0.5 and 1.5 as represented in Eq. (82) and (83).

$$gain' = gain + LSTM\_gain\_out \times 0.5 \times gain \quad \cdots\cdots \text{Eq. (82)}$$

$$loss' = loss + LSTM\_loss\_out \times 0.5 \times loss \quad \cdots\cdots\cdots\cdots\cdots\cdots \text{Eq. (83)}$$

As such, the empirical model used to analyze updated perception effects of covariates on service satisfaction ($Satisfy$) of consumers in the Macaron ride-hailing platform service is presented in Equation (84). The conceptualization of the "Perception Gap" is illustrated in Figure 26.

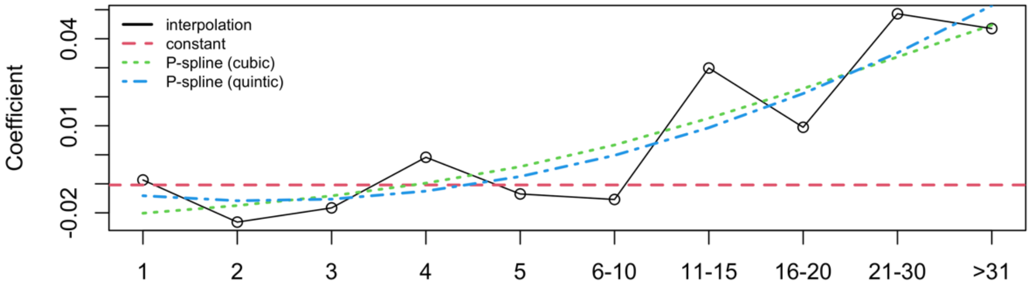
$$\begin{aligned} Satisfy_{i,t} = {} & \beta_0 + \beta_1 Age_i + \beta_2 Gender_i + \beta_3 WaitTime_{i,t} \\ & + \beta_4 PerceivedGainDist_{i,t} + \beta_5 PerceivedLossDist_{i,t} \\ & + \beta_6 PerceivedGainTime_{i,t} + \beta_7 PerceivedLossTime_{i,t} \\ & + \beta_8 PerceivedGainCost_{i,t} + \beta_9 PerceivedLossCost_{i,t} + \epsilon_{it} \cdots\cdots\cdots \text{Eq. (84)} \end{aligned}$$

The estimated results using the updated perceptions are presented in Table 26. Coefficients for covariates as identical to those of described and updated expectations analysis were again fit using both the cubic and quintic P-spline functions as shown in Figure 35. From the figure, it can be seen that the coefficients for all four covariates show similar long-time trends compared to previous analysis. Figure 36 shows the estimated functional relationship as well as the simulated coverage of the 95% Bayesian CIs. The summary of fit of the estimated splines is summarized in Tables 27 and 28.

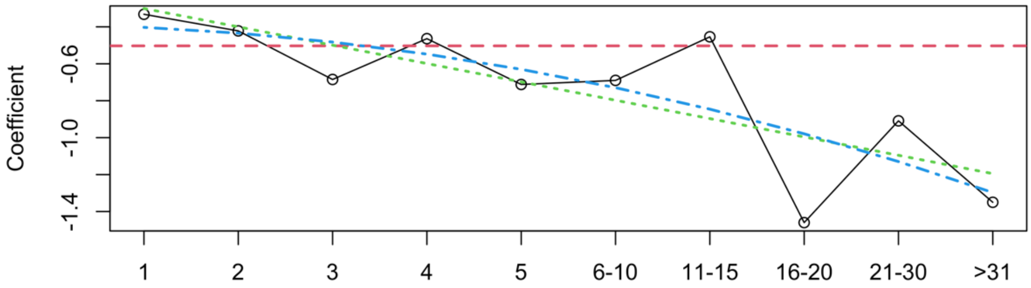
Table 26. Estimation results for service satisfaction with UP

VARIABLES		Cnt=1	Cnt=2	Cnt=3	Cnt=4	Cnt=5	Cnt=6~10	Cnt=11~15
Age		-8.68e-03	-2.32e-02	-1.83e-02	-8.58e-04	-1.35e-02	-1.54e-02	2.99e-02
Gender		-3.32e-01	-4.22e-01	-6.85e-01	-4.64e-01	-7.12e-01	-6.90e-01	-4.54e-01
WaitTime		-3.62e-04	-5.13e-04	3.74e-05	-6.67e-04	-7.13e-04	-4.03e-04	3.20e-04
Distance	Gain	-3.31e-02	-1.34e-01	-1.80e-01	3.24e-02	-6.03e-01	1.37e-01	1.55e-01
	Loss	2.59e-03	-1.76e-02	-1.06e-01	-7.80e-02	-1.80e-01	1.41e-01	4.74e-01
RideTime	Gain	7.40e-04	-5.31e-04	-1.51e-03	-1.66e-03	-1.37e-03	-1.32e-04	-2.07e-03
	Loss	-3.30e-04	3.38e-05	-5.87e-04	-1.10e-03	1.24e-03	-9.26e-04	-1.97e-03
Cost	Gain	4.16e-05	1.10e-04	2.70e-04	7.94e-05	4.49e-04	-3.06e-04	5.40e-04
	Loss	-3.83e-05	5.22e-05	1.56e-05	-4.00e-05	-3.02e-04	-7.88e-05	-3.50e-04
/cut1		-5.8194	-6.1451	-5.5121	-4.0525	-18.3405	-5.3076	-5.5535
/cut2		1.1702	1.6125	2.0011	1.9834	3.0008	2.1368	2.4135
Observations		2,502	1,047	627	468	369	1,146	655
Train Error		0.5544	0.4656	0.4371	0.4923	0.3097	0.3804	0.3512

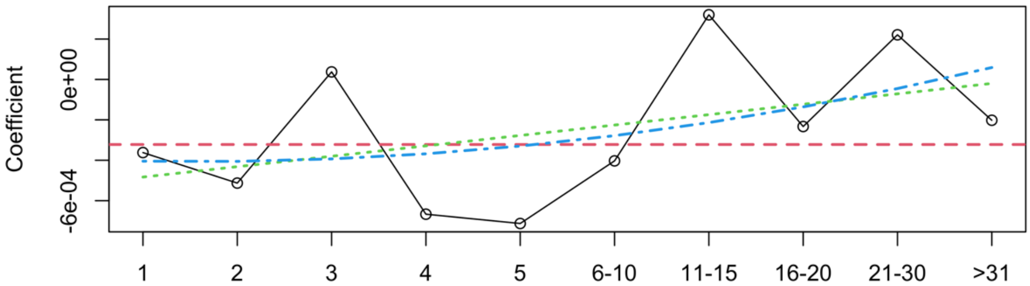
VARIABLES		Cnt=16~20	Cnt=21~30	Cnt>31	ALL
Age		9.48e-03	4.86e-02	4.35e-02	-1.04e-02
Gender		-1.46e+00	-9.09e-01	-1.35e+00	-5.03e-01
WaitTime		-2.33e-04	2.21e-04	-2.02e-04	-3.22e-04
Distance	Gain	-7.65e-02	3.80e-01	-4.47e-01	-2.08e-02
	Loss	2.15e-01	1.92e-01	-2.85e-01	3.14e-03
RideTime	Gain	-8.86e-04	-1.93e-03	1.74e-03	1.21e-04
	Loss	1.26e-04	-2.12e-03	-6.95e-04	-2.64e-04
Cost	Gain	6.22e-04	-6.85e-05	2.45e-04	5.93e-05
	Loss	-1.77e-04	5.93e-05	-7.60e-05	7.16e-06
/cut1		-5.2246	-6.2349	-8.2658	-5.4606
/cut2		2.2326	2.2567	3.0208	1.6896
Observations		417	555	778	8,564
Train Error		0.3701	0.3400	0.2470	0.4574



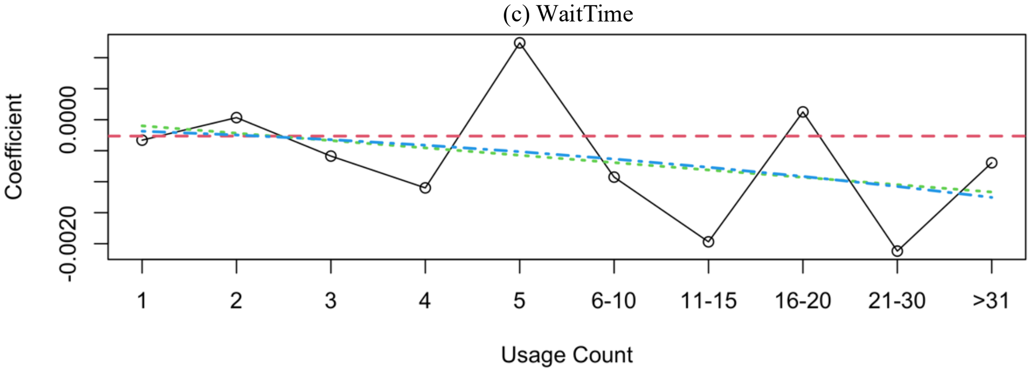
(a) Age



(b) Gender



(c) WaitTime



(d) RideTime_Loss

Figure 35. P-spline fits for satisfaction parameters with perception updating

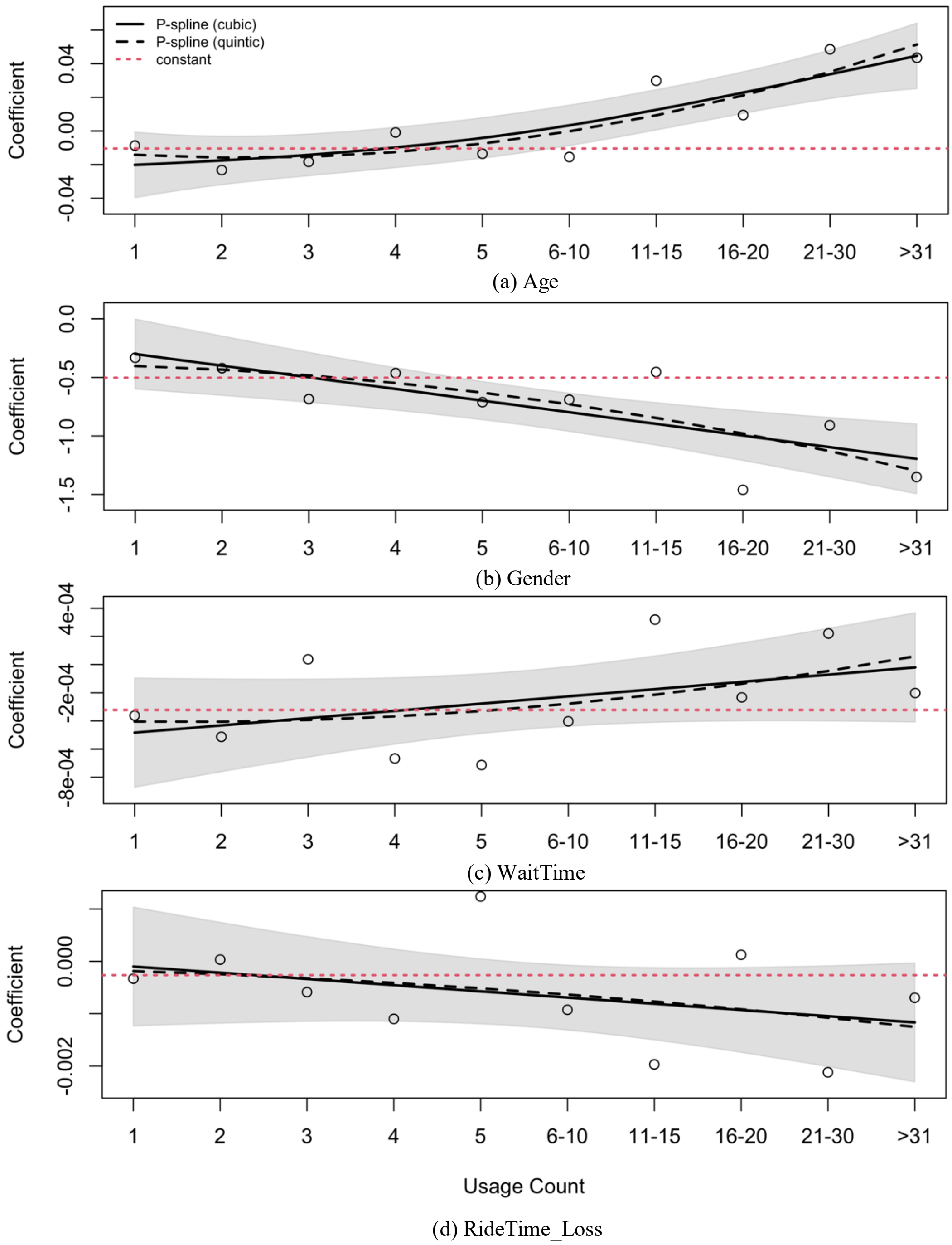


Figure 36. P-spline CIs for satisfaction parameters with perception updating

Table 27. Parametric effects from P-spline fits of satisfaction parameters with UP

		Cubic P-Spline				Quintic P-Spline			
Variable	Parameter	Estimate	Std. Error	t-value	Pr(>$\|t\|$)	Estimate	Std. Error	t-value	Pr(>$\|t\|$)
Age	(Intercept)	4.29e-03	4.45e-03	9.64e-01	3.66e-01	3.47e-03	4.40e-03	7.88e-01	4.56e-01
	x	6.48e-02	1.44e-02	4.49e+00	2.61e-03**	6.55e-02	1.35e-02	4.86e+00	1.84e-03**
	x^2	-	-	-	-	1.82e-01	9.60e-02	1.89e+00	9.94-02.
Gender	(Intercept)	-7.48e-01	8.18e-02	-9.14e+00	1.66e-05***	-7.35e-01	8.61e-02	-8.54e+00	6.00e-05***
	x	-8.94e-01	2.56e-01	-3.49e+00	8.24e-03**	-8.94e-01	2.64e-01	-3.38e+00	1.17e-02*
	x^2	-	-	-	-	-1.38e+00	1.88e+00	-7.34e-01	4.87e-01
WaitTime	(Intercept)	-2.52e-04	1.06e-04	-2.36e+00	4.58e-02*	-2.61e-04	1.15e-04	-2.28e+00	5.68e-02.
	x	4.64e-04	3.33e-04	1.39e+00	2.02e-01	4.63e-04	3.52e-04	1.32e+00	2.30e-01
	x^2	-	-	-	-	1.06e-03	2.51e-03	4.24e-01	6.84e-01
RideTime	(Intercept)	-6.33e-04	3.12e-04	-2.03e+00	7.68e-02.	-6.22e-04	3.40e-04	-1.83e+00	1.10e-01

_Loss	x	-1.07e-03	9.76e-04	-1.09e+00	3.06e-01	-1.07e-03	1.04e-03	-1.02e+00	3.40e-01
	x^2	-	-	-	-	-1.17e-03	7.41e-03	-1.58e-01	8.79e-01

*** p<0.001, ** p<0.01, * p<0.05, . p<0.1

Table 28. Nonparam. effects from P-spline fits of satisfaction parameters with UP

		Cubic P-Spline					Quintic P-Spline				
Variable	Parameter	DF	Sums of Squares	Mean Squares	F-value	Pr(>F)	DF	Sums of Squares	Mean Squares	F-value	Pr(>F)
Age	$f(x)$	7.59e-01	3.87e-04	5.10e-04	2.63e+00	1.47e-01	6.62e-09	3.28e-13	4.95e-05	2.67e-01	6.21e-01
	Residuals	7.24e+00	1.40e-03	1.94e-04	-	-	7.00e+00	1.30e-03	1.86e-04	-	-
Gender	$f(x)$	7.68e-07	2.99e-08	3.89e-02	5.81e-01	4.68e-01	6.62e-09	7.86e-11	1.19e-02	1.67e-01	6.95e-01
	Residuals	8.00e+00	5.36e-01	6.70e-02	-	-	7.00e+00	4.98e-01	7.11e-02	-	-
WaitTime	$f(x)$	7.68e-07	3.63e-14	4.73e-08	4.17e-01	5.37e-01	6.62e-09	6.95e-16	1.05e-07	8.32e-01	3.92e-01
	Residuals	8.00e+00	9.07e-07	1.13e-07	-	-	7.00e+00	8.84e-07	1.26e-07	-	-

RideTime _Loss	$f(x)$	7.68e-07	6.72e-14	8.75e-08	9.01e-02	7.72e-01	6.62e-09	1.48e-15	2.24e-07	2.03e-01	6.66e-01
	Residuals	8.00e+00	7.77e-06	9.71e-07	-	-	7.00e+00	7.74e-06	1.11e-06	-	-

*** p<0.001, ** p<0.01, * p<0.05, . p<0.1

5.4.2.4 Comparison of Models

This section compares the analysis results presented in the previous sections. Figures 37 to 40 illustrates both the interpolation of time-varying coefficients estimated from models with described expectations, updated expectations, and updated perceptions, as well as their P-spline fits. It can be easily seen that the overall time-varying characteristics of the estimated coefficients of the three models are very similar. At the same time, it can be seen that the influence of the attributes on consumers' service satisfaction changes differently across the models.

Again, the graphs show estimated coefficients with usage count τ on the horizontal axis and the estimated coefficients for the covariate effects on the vertical axis. As mentioned in Section 5.4.2.1, the general time-varying trend in the age coefficients was that the younger the consumer, the higher the satisfaction at first use of the service. The influences of young age on service satisfaction were similar in all three models at this point in the usage experience. Nonetheless, after the 2nd time of service use, the increase in satisfaction among younger consumers declined more steeply in the models that accounted for the expectations effect than in the models with the updated perceptions effect. After the 5th time of use, however, the direction of the influence completely reversed in all three models, i.e., the older the user, the higher the satisfaction with the service. Thereafter, the influence of older age on increased satisfaction increased more steeply in the updated perceptions model. The magnitude of the increase in the influence of age covariate was greater in the order of the updated perceptions, updated expectations, and described expectations models. Overall, the sensitivity of age to user satisfaction was greatest in the updated perceptions model and least in the described expectations model.

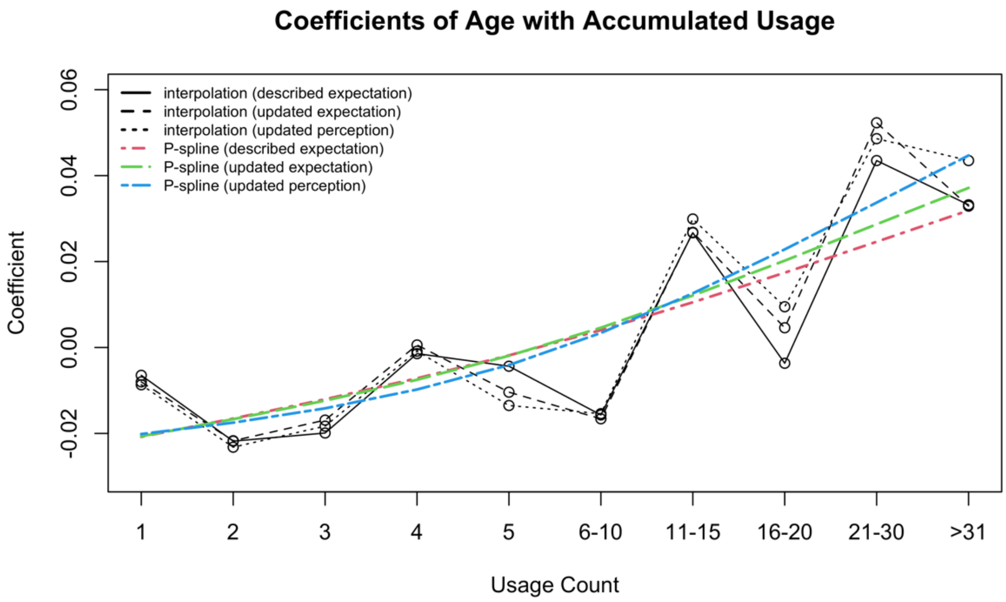


Figure 37. P-spline fits comparison for Age coefficients

Overall, satisfaction with the service was higher among men than among women, and this tendency was more heavily influenced as the number of uses increased. For the first use of the service, the influence of gender was greatest for the model incorporating updated perceptions effect and least for described expectations effect. However, the degree of increase in the influence of gender was lowest in the updated perception model, so that the final influence corresponding to the effect of accumulated experience was ultimately lowest. In contrast, the model with the described expectations with steepest increase had the highest degree of influence in the end. Comparison of the models showed that the model with the described expectations was the most sensitive overall and the model with the updated perceptions was the least sensitive to the covariate gender.

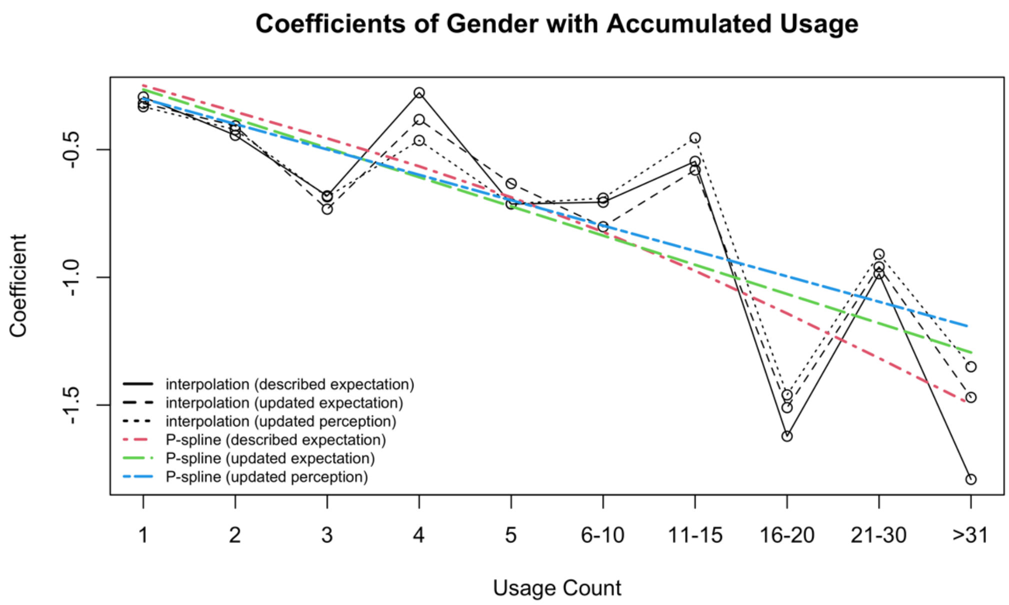


Figure 38. P-spline fits comparison for Gender coefficients

Trends in changes in wait time coefficients were generally similar across models, with a similar degree of decline in their effects on service satisfaction. Nevertheless, the model with the updated perceptions was the most sensitive and the one with the updated expectations was the least sensitive to the number of times the service was used. While the model with updated expectations showed the smallest impact of wait time on service satisfaction at first use, the model with updated perceptions, which once showed the largest impact at first use, showed the smallest impact at cumulative use.

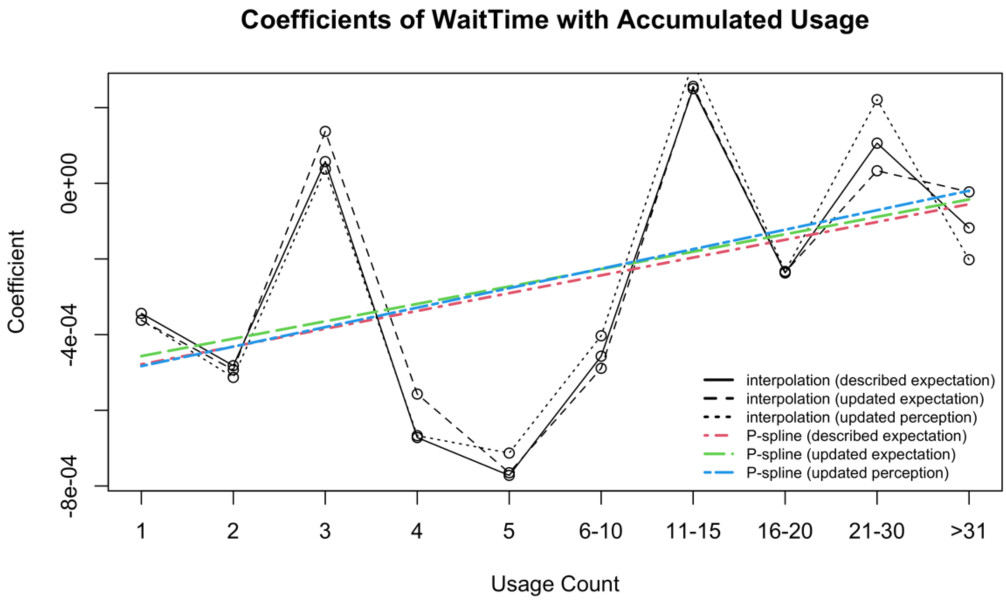


Figure 39. P-spline fits comparison for WaitTime coefficients

It can also be seen from Figure 41 that the time-varying trends of $RideTime\_Loss$ coefficients differ significantly across models. At first use, the model with described expectations revealed almost zero influence of loss in ride time on service satisfaction; however, with accumulated usage experience, reduction in such loss led to marginal increase in service satisfaction. While reduction of loss in ride time led to increased service satisfaction in models with updated expectations and updated perceptions, the degree of increase in the level of impact was greater in the model with updated perception with accumulated use. Integration of updated perception yielded significant impact of reduction in $RideTime\_Loss$ to increased service satisfaction with increased number of use of the service.

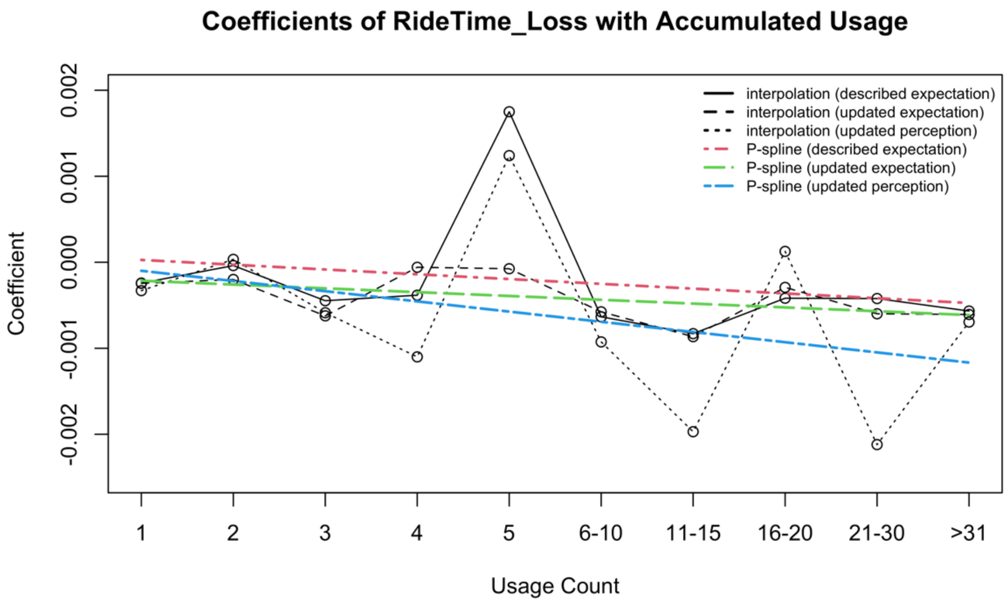


Figure 40. P-spline fits comparison for RideTime_Loss coefficients

Finally, the prediction errors of the three models were compared to determine which model performed better in explaining the data. While optimizers are algorithms or methods used to change neural network attributes such as weights and learning rate in order to reduce losses, the loss function in a neural network quantifies the difference between the expected outcome and the outcome produced by the deep learning model. In other words, it is used to quantify how well or poorly the model performs. The greater the loss (train error), the greater the error in the prediction.

Figure 41 illustrates the change in prediction error for models that incorporate described expectations, updated expectations, and updated perceptions. The two models with described and updated expectations reveal identical prediction loss at first use (0.5563), with updated preference model performing slightly better (0.5544). However, as the consumers' number of service use increases, the prediction error for each model decreases at a different rate. Observing that the prediction error for described expectations model decreases, it can be speculated that the service itself optimizes the notification function as to minimize its error with actual delivery of service for users. But however well the system learns through algorithm to notify the messages that form described expectations with least error, users themselves update their expectations on error, and in turn on service components, with accumulated experience. This is well observed in Figure 41, where the decrease in prediction error for updated expectations model is steeper than that with described expectations.

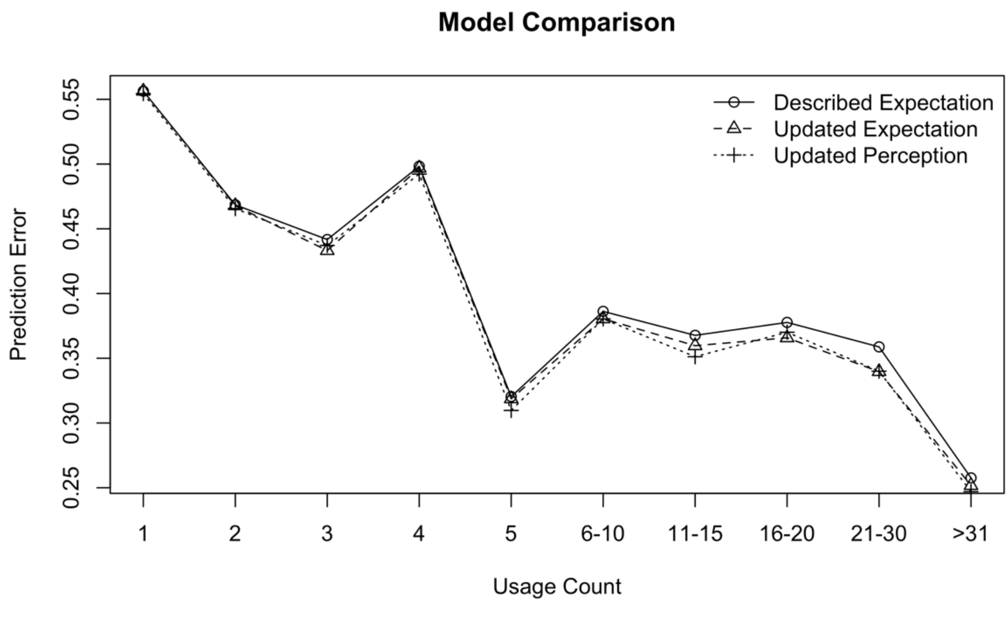


Figure 41. Prediction error comparison

Finally, the graph confirms our assumption that people update their perceptions of expected gains and losses with respect to service use. While the prediction error for the model with updated perceptions decreases with repeated use, the error reduces at a higher level than for other models. The final prediction errors for described expectations, updated expectations, and the updated perceptions models are 0.2576, 0.2517, and 0.2470, respectively. Lower prediction error indicates better explanatory power of the model when incorporating the time-varying behavior of platform users.

5.4.3 Stream-of-Time Effects (by Times of the Day)

In this section, the stream-of-time effects of covariates in explaining consumers' use of ride-hailing platforms are analyzed using models based on P-splines, a semiparametric approach. The flow of time is observed by the times of the day. Analysis that observes stream-of-time effects by the days of the week can be found in the Appendix. By identifying the stream-of-time effects, the seasonality of consumer behaviors can be observed.

The data pooled from Macaron Taxi totaling 8,564 successful immediate rides were categorized by the times of the day to analyze stream-of-time effects by usage hours. The number of successful rides as categorized by times are presented in Figure 42. From the figure, it is observed that the usage of the service most occur during the morning hours (7am to 1pm), followed by 6pm and 10pm. This coincides with the trend reported in the 2019 Kakao Mobility Report that states that its peak usage hours were (1) 7am to 10am, (2) 5pm to 7pm, and (3) 9pm to 2am in year 2019, with the greatest number of usages at 8am (ref). As such, while Kakao Taxi is the largest mobility service company in South Korea, Macaron is also deemed as a representative sample as of usage frequency trends of ride-hailing platforms by times of the day.

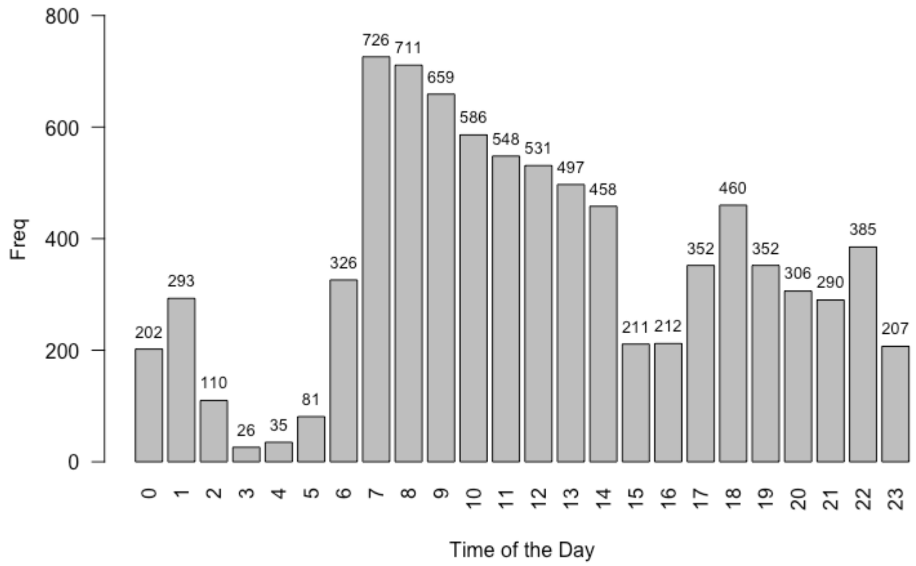


Figure 42. Service usage frequency by times of the day

To investigate the usage interval and the total number of usage of passengers in the Macaron ride-hailing platform service, negative binomial regression was used alike the general model; however, only the model without the interaction effects of travel distance and speed (Model 1) was utilized. Again, our interest is to identify the influence of gains and losses of specified covariates in the interval of consumers' use of service as well as their total number of usages. For usage interval ($IntvlUse$) analysis, consumers with only a single use of service were excluded. The estimation results for usage interval are presented in Table 29 below. The significant variables for each time are again summarized in Table 30 for clarity.

It is notable that for times of 3AM and 4AM, estimation did not converge due to the lack of sample size. The results also reveal that coefficients for covariates of Age, $Gender$, and $Satisfy_{t-1}$, as well as the constant, were consistently significant on most times of the day. Covariates with regards to distance, ride time, and cost were almost always insignificant (≤ 6). The estimated paths for coefficients of covariates that were consistently significant and varies by the times of the day were realized using both the cubic and quintic P-spline functions as shown in Figure 43. It depicts estimated coefficient with calendar time τ in hours on the horizontal axis and estimated coefficients for covariate effects and constant term on the vertical axis. The solid line represents full interpolation of estimated coefficients, and the constant coefficient estimated in the generic model is also depicted.

Table 29. Estimation results for usage interval by times

VARIABLES		0AM	1AM	2AM	3AM	4AM	5AM	6AM	7AM
Age		**0.0727****	**-0.0894*****	**0.2200*****	-	-	**0.1180*****	0.0044	**-0.0283*****
		(0.0297)	**(0.0206)**	**(0.0552)**	-	-	**(0.0425)**	(0.0123)	**(0.0067)**
Gender		**1.1350*****	**1.5050*****	**-3.4070*****	-	-	1.4420	**0.7320*****	**-0.7930*****
		(0.4340)	**(0.3610)**	**(0.6970)**	-	-	(0.9490)	**(0.2520)**	**(0.1610)**
WaitTime $_{t-1}$		-0.0006	**0.0006*****	0.0005	-	-	**0.0035****	-0.0003	0.0005
		(0.0006)	**(0.0002)**	(0.0008)	-	-	**(0.0016)**	(0.0006)	(0.0003)
Distance	Gain $_{t-1}$	0.1780	**0.0979***	-0.1070	-	-	-0.3830	0.1520	0.0234
		(0.2250)	**(0.0500)**	(0.1300)	-	-	(0.4200)	(0.1670)	(0.0404)
	Loss $_{t-1}$	-0.3860	0.0479	**-0.2880***	-	-	**-1.8530***	0.1390	0.0243
		(0.2880)	(0.0618)	**(0.1750)**	-	-	**(0.9510)**	(0.1140)	(0.0199)
RideTime	Gain $_{t-1}$	**-0.0017***	3.88e-05	-0.0004	-	-	0.0015	0.0006	0.0003
		(0.0009)	(0.0005)	(0.0010)	-	-	(0.0022)	(0.0009)	(0.0005)
	Loss $_{t-1}$	0.0007	-0.0002	0.0010	-	-	-0.1390	0.0012	0.0005
		(0.0023)	(0.0005)	(0.0014)	-	-	(2.087e+08)	(0.0012)	(0.0005)
Cost	Gain $_{t-1}$	-0.0004	**-0.0002***	-0.0012	-	-	-0.0008	-0.0006	-0.0001
		(0.0003)	**(8.88e-05)**	(0.0011)	-	-	(0.0006)	(0.0004)	(0.0001)
	Loss $_{t-1}$	6.77e-05	**0.0005*****	-0.0004	-	-	-0.0004	-0.0002	2.16e-05
		(0.0003)	**(8.96e-05)**	(0.0003)	-	-	(0.0006)	(0.0003)	(0.0001)
Satisfy $_{t-1}$		**-1.1330***	**-0.5290****	-0.2590	-	-	-1.6840	**-1.078****	**0.5460*****

		(0.6650)	**(0.2380)**	(0.3940)	-	-	(1.4650)	**(0.4290)**	**(0.1950)**
Constant		-0.0122	**4.0780*****	**-4.8170*****	-	-	**-4.5460****	**1.7610*****	**3.2170*****
		(1.125)	**(0.7430)**	**(1.7890)**	-	-	**(2.3190)**	**(0.6180)**	**(0.3560)**
Observations		75	205	64			14	110	438
/lnalpha		-0.1200	**-0.2080***	**-0.4450***	-	-	**-1.2220****	-0.1080	**0.2100*****
		(0.1790)	**(0.1210)**	**(0.2390)**	-	-	**(0.5660)**	(0.1440)	**(0.0707)**
Log-Likelihood		-208.8408	-520.4036	-165.5782	-	-	-42.0036	-340.4732	-1206.2964

VARIABLES		8AM	9AM	10AM	11AM	12PM	13PM	14PM	15PM
Age		**-0.0302*****	-0.0041	**0.0448*****	**-0.0302****	**0.0252***	**-0.0275****	-0.0052	0.0204
		(0.0057)	(0.0094)	**(0.0133)**	**(0.0128)**	**(0.0136)**	**(0.0139)**	(0.0140)	(0.0464)
Gender		0.0802	**0.4360*****	0.2770	0.4610	0.6090	**-0.9820***	**-0.9500*****	0.3980
		(0.1320)	**(0.1580)**	(0.2570)	(0.3840)	(0.3820)	**(0.5020)**	**(0.2480)**	(0.4960)
$WaitTime_{t-1}$		-6.73e-06	4.72e-05	**-0.0009****	**-0.0005****	0.0005	0.0003	0.0001	0.0007
		(0.0002)	(0.0002)	**(0.0004)**	**(0.0002)**	(0.0005)	(0.0005)	(0.0003)	(0.0006)
Distance	$Gain_{t-1}$	-0.0273	0.0352	0.0203	-0.0465	-0.0519	-6.34e-05	**-0.2520***	-0.2410
		(0.0622)	(0.1070)	(0.1440)	(0.0845)	(0.1020)	(0.4290)	**(0.1420)**	(0.2570)
	$Loss_{t-1}$	-0.0140	-0.1660	0.0178	-0.0948	**0.0380****	-0.0899	0.4300	**-0.0229***
		(0.0172)	(0.1160)	(0.0347)	(0.0738)	**(0.0169)**	(0.1820)	(0.2780)	**(0.0134)**
RideTime	$Gain_{t-1}$	-0.0003	0.0005	-0.0002	0.0009	**0.0016***	0.0014	0.0004	9.82e-05
		(0.0004)	(0.0006)	(0.0006)	(0.0006)	**(0.0009)**	(0.0012)	(0.0007)	(0.0011)

	Loss $_{t-1}$	0.0002	-5.95e-05	-0.0001	**-0.0009***	-2.85e-05	0.0012	0.0013	9.24e-05
		(0.0004)	(0.0006)	(0.0002)	**(0.0005)**	(0.0009)	(0.0010)	(0.0009)	(0.0013)
Cost	Gain $_{t-1}$	**-0.0004*****	**-0.0003***	0.0003	5.38e-05	-0.0002	-0.0001	**-0.0005****	**0.0009***
		(0.0001)	**(0.0002)**	(0.0003)	(0.0003)	(0.0003)	(0.0004)	**(0.0002)**	**(0.0005)**
	Loss $_{t-1}$	-5.86e-05	8.45e-05	**0.0002****	0.0002	-0.0001	-0.0002	-0.0003	7.03e-05
		(8.32e-05)	(0.0001)	**(0.0001)**	(0.0002)	(0.0003)	(0.0002)	(0.0003)	(0.0003)
Satisfy $_{t-1}$		**0.5960*****	**0.4330***	**-1.0560*****	0.5010	-0.1150	**-1.3160****	-0.4180	**0.9670*****
		(0.1520)	**(0.2340)**	**(0.3300)**	(0.3260)	(0.3730)	**(0.5780)**	(0.2720)	**(0.3060)**
Constant		**2.8520*****	**2.1710*****	0.8030	**3.4030*****	0.5080	**4.3130*****	**3.0460*****	0.5250
		(0.2400)	**(0.4430)**	(0.6580)	**(0.6230)**	(0.5950)	**(0.9500)**	**(0.6390)**	(2.1810)
Observations		377	249	126	116	90	83	117	45
/lnalpha		0.0416	**0.1990****	-0.1290	0.1020	-0.2620	-0.0417	**-0.3650****	**-0.7870*****
		(0.0803)	**(0.0902)**	(0.1270)	(0.1380)	(0.1640)	(0.1660)	**(0.1530)**	**(0.2420)**
Log-Likelihood		-987.9757	-817.8342	-455.9177	-386.5990	-292.3635	-289.2689	-353.1928	-145.9683

VARIABLES		16PM	17PM	18PM	19PM	20PM	21PM	22PM	23PM
Age		-0.0862	0.0069	**0.0249****	**-0.0271***	-0.0040	0.0264	**0.0163****	0.0066
		(0.0658)	(0.0153)	**(0.0122)**	**(0.0151)**	(0.0072)	(0.0210)	**(0.0080)**	(0.0147)
Gender		-0.4110	**-1.2010*****	0.3270	**0.7340*****	-0.0435	-0.0592	-0.2630	**-1.1550*****
		(0.9650)	**(0.3750)**	(0.2100)	**(0.2620)**	(0.2850)	(0.610)	(0.1890)	**(0.2910)**
WaitTime $_{t-1}$		**-0.0020***	-0.0002	0.0003	-0.0005	-0.0006	**-0.0009****	-7.28e-05	-0.0001
		(0.0011)	(0.0002)	(0.0003)	(0.0004)	(0.0005)	**(0.0004)**	(0.0003)	(0.0005)

Distance	Gain $_{t-1}$	-0.2030	0.1120	-0.0227	0.0511	0.0312	0.1380	-0.0206	-0.0452
		(0.3110)	(0.1600)	(0.0993)	(0.0970)	(0.0442)	(0.1540)	(0.0514)	(0.0570)
	Loss $_{t-1}$	**-0.5500***	-0.2480	-0.2490	0.4120	0.1550	0.1650	**-0.3190*****	-0.0421
		(0.2920)	(0.2540)	(0.2660)	(0.3400)	(0.1120)	(0.1850)	**(0.1200)**	(0.2090)
RideTime	Gain $_{t-1}$	0.0003	**0.0015***	**0.0012***	**0.0031*****	-0.0006	-0.0013	0.0006	**0.0027*****
		(0.0027)	**(0.0008)**	**(0.0007)**	**(0.0008)**	(0.0006)	(0.0015)	(0.0007)	**(0.0010)**
	Loss $_{t-1}$	-0.0009	-0.0009	0.0009	0.0016	**0.0018*****	**0.0049*****	-0.0007	0.0002
		(0.0013)	(0.0008)	(0.0009)	(0.0011)	**(0.0007)**	**(0.0017)**	(0.0005)	(0.0012)
Cost	Gain $_{t-1}$	0.0015	3.37e-05	**-0.0004*****	-6.98e-05	7.93e-05	0.0020	-9.08e-06	-5.24e-05
		(0.0012)	(0.0002)	**(0.0001)**	(0.0002)	(0.0003)	(0.0014)	(0.0001)	(8.50e-05)
	Loss $_{t-1}$	0.0004	**0.0005****	-0.0002	-0.0004	**-0.0003***	**-0.0005*****	-3.32e-05	0.0001
		(0.0003)	**(0.0002)**	(0.0003)	(0.0003)	**(0.0002)**	**(0.0002)**	(5.00e-05)	(0.0003)
Satisfy $_{t-1}$		-0.4430	**-0.7310***	0.6270	-0.5710	0.2830	**-1.7460*****	-0.4260	0.1520
		(0.8790)	**(0.4100)**	(0.3860)	(0.4470)	(0.7470)	**(0.5820)**	(0.2790)	(0.5520)
Constant		**6.8440***	**2.7110*****	**1.2350*****	**2.1940*****	**2.5570*****	1.9520	**1.6430*****	**2.6330*****
		(3.7280)	**(0.6470)**	**(0.4680)**	**(0.5060)**	**(0.470)**	(1.1870)	**(0.3870)**	**(0.6420)**
Observations		26	121	130	98	80	55	161	42
/lnalpha		**-1.3840*****	-0.0642	0.0100	**-0.3620****	**-0.3480***	**0.6620*****	0.0059	**-0.9860*****
		(0.4040)	(0.1330)	(0.1290)	**(0.1690)**	**(0.1800)**	**(0.2000)**	(0.1200)	**(0.2660)**
Log-Likelihood		-78.4276	-388.8822	-441.0752	-282.4286	-251.1285	-160.5067	-482.6943	-135.8860

*** p<0.01, ** p<0.05, * p<0.1; Standard errors are in parenthesis

Table 30. Summary of significant variables for usage interval by times

VARIABLES		0AM	1AM	2AM	3AM	4AM	5AM	6AM	7AM
Age		(+)	(-)	(+)			(+)		(-)
Gender		(+)	(+)	(-)				(+)	(-)
WaitTime $_{t-1}$			(+)				(+)		
Distance	Gain $_{t-1}$		(+)						
	Loss $_{t-1}$			(-)			(-)		
RideTime	Gain $_{t-1}$	(-)							
	Loss $_{t-1}$								
Cost	Gain $_{t-1}$		(-)						
	Loss $_{t-1}$		(+)						
Satisfy $_{t-1}$		(-)	(-)					(-)	(-)
Constant			(+)	(-)			(-)	(+)	(+)

VARIABLES		8AM	9AM	10AM	11AM	12PM	13PM	14PM	15PM
Age		(-)		(+)	(-)	(+)	(-)		
Gender			(+)				(-)	(-)	
WaitTime $_{t-1}$				(-)	(-)				
Distance	Gain $_{t-1}$							(-)	
	Loss $_{t-1}$					(+)			(-)

RideTime	Gain $_{t-1}$					(+)			
	Loss $_{t-1}$				(-)				
Cost	Gain $_{t-1}$	(-)	(-)					(-)	(+)
	Loss $_{t-1}$			(+)					
Satisfy $_{t-1}$		(+)	(+)	(-)			(-)		(+)
Constant		(+)	(+)		(+)		(+)	(+)	

VARIABLES		16PM	17PM	18PM	19PM	20PM	21PM	22PM	23PM
Age				(+)	(-)			(+)	
Gender			(-)		(+)				(-)
WaitTime $_{t-1}$		(-)					(-)		
Distance	Gain $_{t-1}$								
	Loss $_{t-1}$	(-)						(-)	
RideTime	Gain $_{t-1}$		(+)	(+)	(+)				(+)
	Loss $_{t-1}$					(+)	(+)		
Cost	Gain $_{t-1}$			(-)					
	Loss $_{t-1}$		(+)			(-)	(-)		
Satisfy $_{t-1}$			(-)				(-)		
Constant		(+)	(+)	(+)	(+)	(+)		(+)	(+)

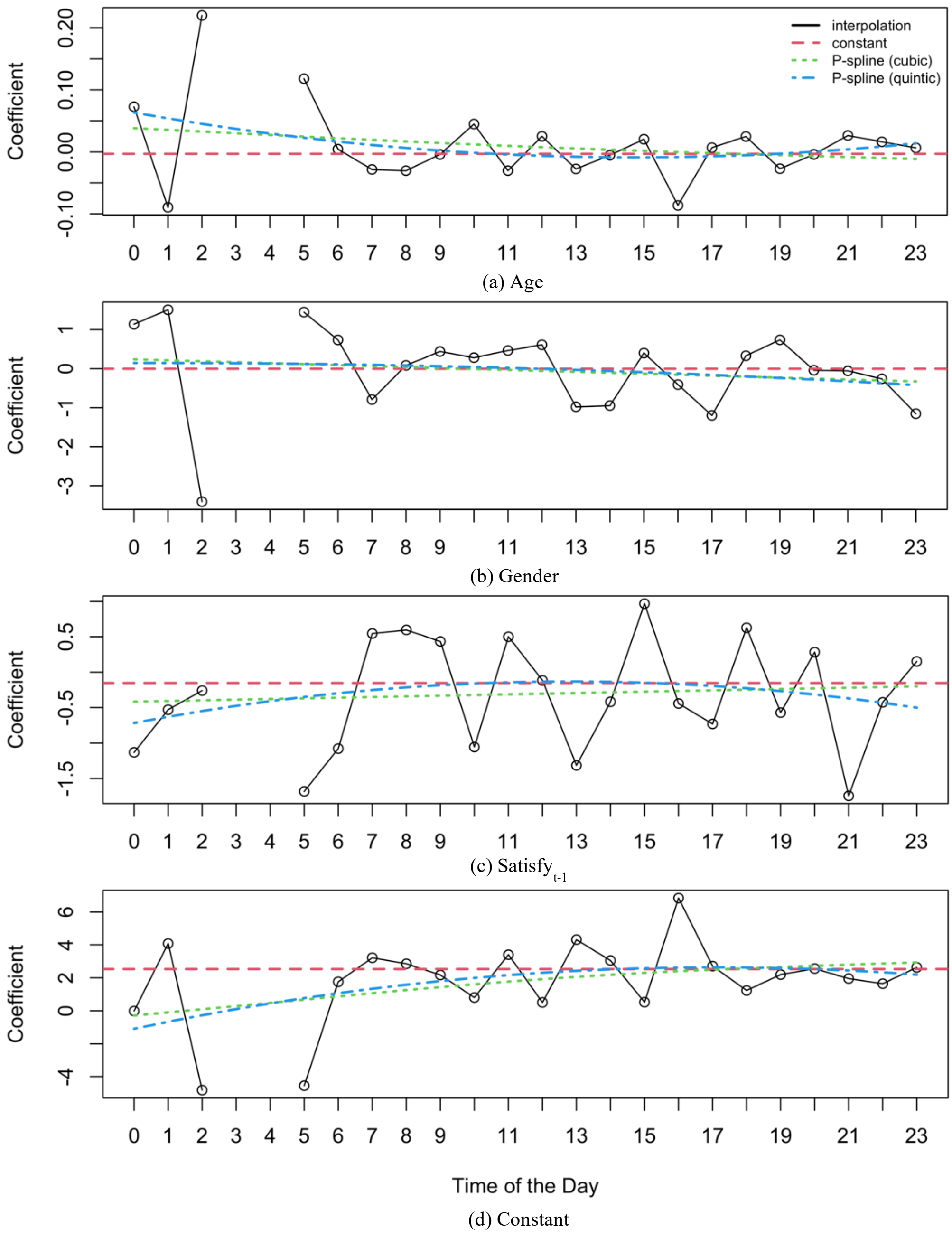


Figure 43. P-spline fits for Usage Interval parameters by times of the day

From the figure, we obtain rather smooth curves for all covariates suggesting long-term trends in their time-varying influence on usage interval. Amongst, the coefficients for $Gender$ varies the least, almost coinciding with the constant coefficient from the generic model. The result suggests that while women have a slightly larger usage interval until around noon, men have a slightly larger usage interval afterwards. Overall, the effect of the consumer's gender on the usage interval of service is rather small.

It is also revealed that the older the person the greater the usage interval of the ride-hailing platform, meaning that younger people use the service more often. However, such influence of age on the total service use decreases by time. The influence of age on usage interval is the least around 6pm. To add, the results also show that until around 11AM, being women have a larger usage interval, and then afterwards, being men yield a larger usage interval. This suggests that men use the service more frequently until around noon, and women use the service more frequently in the afternoons and evenings.

Lastly, the results show that the lesser satisfied the consumers in their previous use of service, the greater the usage interval, suggesting that greater satisfaction leads to more frequent usage of the service. Such influence of previous user satisfaction on usage interval is at is greatest around midnight (0AM~2AM), and as the time flows, the level of influence decreases.

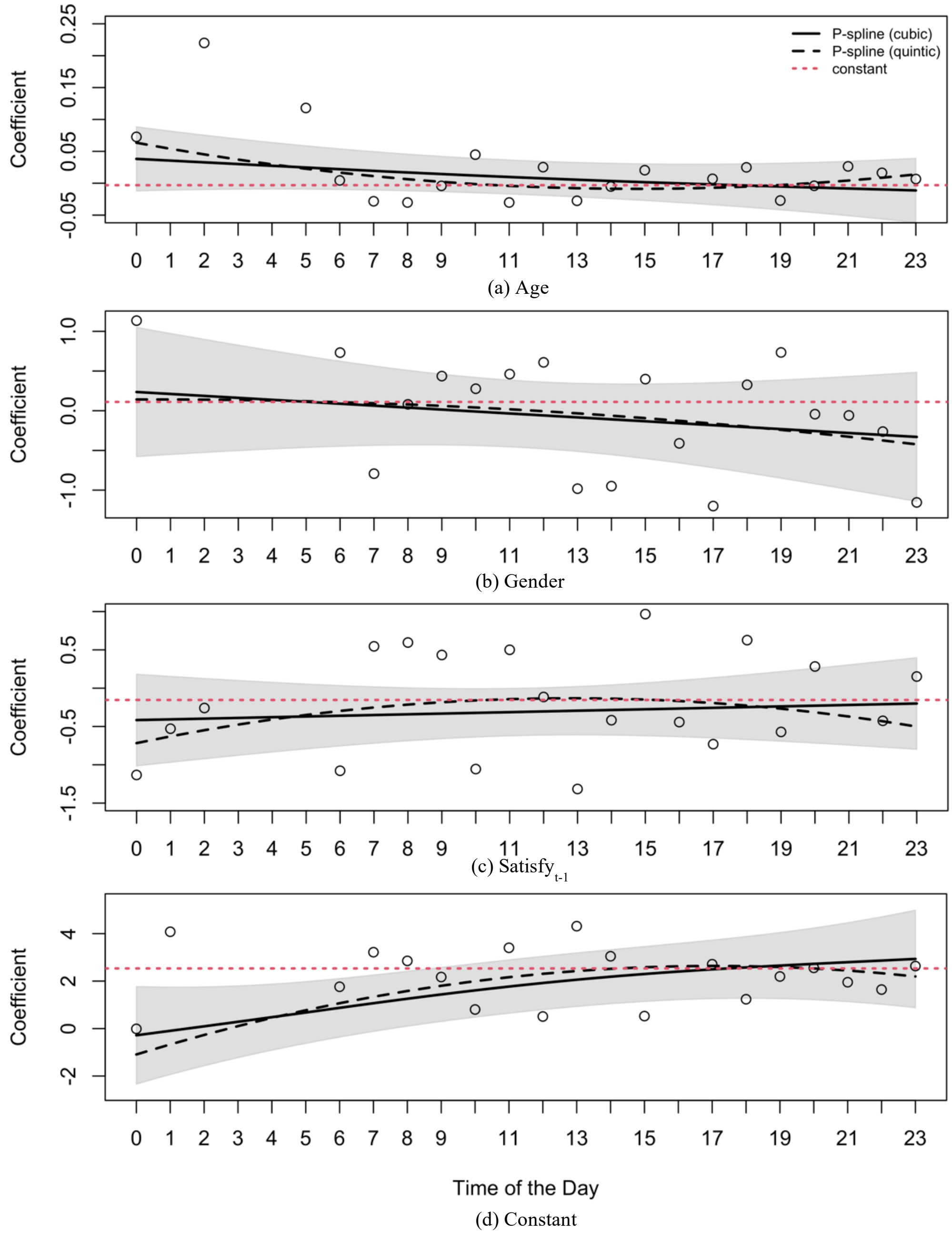


Figure 44. P-spline CIs for Usage Interval parameters by times of the day

The coverage of the 95% Bayesian confidence interval (CI) for each smoothing was calculated using Eq. (44). Figure 44 shows the estimated functional relationship as well as the simulated coverage of the 95% Bayesian CIs. To add, the summary of fit of the estimated splines is summarized in Table 31 and 32, each showing the approximate significance of parametric and nonparametric effects, respectively.

Putting the results together, it is noticeable that for covariates of Age, $Gender$, and $Satisfy_{t-1}$, the coefficients that are estimated from the generic model completely falls within the Bayesian CIs of the cubic spline. This indicates that all models sufficiently explain the data, with the nonparametric model showing the long-term variance of time. Likewise, because all nonparametric effects of the covariates are insignificant, this indicates that that the linear effects need to be considered. There are no short-term trends in the effect of covariates on the usage interval of ride-hailing platform's customers, and their preferences are rather consistent with time.

Meanwhile, the constant coefficients that is estimated from the generic model do not fall completely within the Bayesian CIs of the cubic spline estimations for the constant, suggesting that the nonparametric model should be preferred in fitting the coefficients of the constant term that varies with time. Nonetheless, the nonparametric effects are insignificant, indicating that the linear effects involved in the smooth need to be considered, again representing long-term trend of time-varying coefficients.

Table 31. Parametric effects from P-spline fits of usage intvl parameters by times

Variables	Parameter	Cubic P-Spline				Quintic P-Spline			
		Estimate	Std. Error	t-value	Pr(>|t|)	Estimate	Std. Error	t-value	Pr(>|t|)
Age	(Intercept)	0.0105	0.0126	0.83	0.4137	0.0093	0.0126	0.74	0.4665
	x	-0.0495	0.0421	-1.18	0.2524	-0.0497	0.0416	-1.19	0.2456
	x^2	-	-	-	-	0.3518	0.3100	1.14	0.2691
Gender	(Intercept)	-0.0470	0.2141	-0.22	0.8282	-0.0428	0.2198	-0.19	0.8476
	x	-0.5657	0.7114	-0.80	0.4350	-0.5657	0.7273	-0.78	0.4453
	x^2	-	-	-	-	-1.1779	5.4178	-0.22	0.8300
$Satisfy_{t-1}$	(Intercept)	-0.3083	0.1575	-1.96	0.0630.	-0.2947	0.1583	-1.86	0.0768.
	x	0.2162	0.5232	0.41	0.6835	0.2162	0.5240	0.41	0.6841
	x^2	-	-	-	-	-3.7723	3.9030	-0.97	0.3448
Constant	(Intercept)	1.6420	0.4733	3.47	0.0022**	1.6770	0.4769	3.52	0.0021**
	x	3.2170	1.6153	1.99	0.0593.	3.2870	1.5783	2.08	0.0497*
	x^2	-	-	-	-	-13.5010	11.7567	-1.15	0.2637

*** $p<0.001$, ** $p<0.01$, * $p<0.05$, . $p<0.1$

Table 32. Nonparam. effects from P-spline fits of usage intvl parameters by times

Variables	Parameter	Cubic P-Spline					Quintic P-Spline				
		DF	Sum Sq.	Mean Sq.	F-value	Pr(>F)	DF	Sum Sq.	Mean Sq.	F-value	Pr(>F)
Age	$f(x)$	0.1307	5.629e-04	0.0043	1.14	0.2978	5.34e-09	2.31e-12	4.326e-04	0.11	0.7380
	Residuals	21.8693	0.0828	0.0038	-	-	2.10e+01	7.904e-02	3.764e-03	-	-
Gender	$f(x)$	6.562e-07	6.059e-08	0.0923	0.08	0.7748	5.34e-09	7.373e-10	0.1381	0.12	0.7324
	Residuals	2.200e+01	2.420e+01	1.1002	-	-	2.10e+01	2.415e+01	1.1500	-	-
$Satisfy_{t-1}$	$f(x)$	6.562e-07	3.222e-07	0.4910	0.83	0.3735	5.34e-09	4.406e-10	0.0825	0.14	0.7137
	Residuals	2.200e+01	1.309e+01	0.5950	-	-	2.10e+01	1.253e+01	0.5968	-	-
Constant	$f(x)$	0.3681	2.7540	7.4830	1.39	0.2505	5.34e-09	2.310e-08	4.3260	0.80	0.3816
	Residuals	21.6319	116.0820	5.3660	-	-	2.10e+01	1.137e+02	5.4150	-	-

*** p<0.001, ** p<0.01, * p<0.05, . p<0.1

Likewise, the same empirical model from the generic model was used to observe stream-of-time effects of the total service use ($NbrUse$) of consumer n at occasion t. However, only the model without the interaction effects of travel distance and speed (Model 1) were utilized. Our interest in using Model 1 is to identify the influence of gains and losses of specified covariates on consumers' total numbers of usages. The estimation results are presented in Table 33 below, and the significant variables for each day are again summarized in Table 34 for clarity.

The results reveal that the coefficients for covariates of Age and $Gender$ were consistently significant at most times of the day, followed by covariates of $Saisfy$, which was almost always significant other than 1AM to 5AM, and $WaitTime$. Coefficients for covariates that were consistently significant and varies by day of the week were fit using both the cubic and quintic P-spline functions as shown in Figure 45.

From the figure, it can be seen that the coefficients for all four covariates show long-time trends. Meanwhile, the results show that the older the person, and the lesser the wait time, the greater the total service use of ride-hailing platform, and their level of influence on the total service use decrease slowly as the time flows. The quintic spline for wait time additionally shows that the level of influence in reduced wait time may temporarily diminish until about noon and increase back afterwards. Overall, this suggests that the effect of wait time reduction on the total service use is most effective during night hours. Nonetheless, coefficients for these two covariates vary less with respect to time than those of other covariates.

Table 33. Estimation results for total usage by times

VARIABLES		0AM	1AM	2AM	3AM	4AM	5AM	6AM	7AM
Age		**-0.0507*** **	**0.1264*** **	**0.0756** **	-0.0052	**0.0351** **	**0.0311*** **	**-0.0149** **	**0.0528*** **
		(0.0098)	**(0.0086)**	**(0.0307)**	(0.0282)	**(0.0143)**	**(0.0079)**	**(0.0072)**	**(0.0066)**
Gender		**-0.6929*** **	**-1.1703*** **	**-1.064** **	-0.1067	-0.1734	0.1150	**-1.2118*** **	**0.9964*** **
		(0.1600)	**(0.1777)**	**(0.4180)**	(0.5247)	(0.3868)	(0.1997)	**(0.1365)**	**(0.1033)**
WaitTime		6.06e-05	-0.0002	**-0.0013*** **	-5.55e-05	-6.30e-05	0.0005	-5.92e-05	**-0.0007*** **
		(0.0002)	(0.0002)	**(0.0004)**	(0.0010)	(0.0009)	(0.0004)	(0.0003)	**(0.0003)**
Distance	Gain	-0.0838	**0.0585** **	-0.0413	0.0340	0.0334	-0.0195	**-0.0675***	0.0446
		(0.0581)	**(0.0276)**	(0.0736)	(0.0815)	(0.0932)	(0.0533)	**(0.0401)**	(0.0410)
	Loss	0.1137	-0.0113	0.0055	0.0485	0.1211	0.0423	0.0434	0.0378
		(0.1040)	(0.0361)	(0.0700)	(0.1840)	(0.2227)	(0.1353)	(0.0284)	(0.0292)
RideTime	Gain	**0.0018*** **	**-0.0009*** **	0.0001	-0.0008	-4.03e-05	0.0003	**-0.0014*** **	-0.0003
		(0.0004)	**(0.0003)**	(0.0007)	(0.0017)	(0.0008)	(0.0006)	**(0.0004)**	(0.0003)
	Loss	-0.0008	0.0001	-0.0003	-0.0036	-0.0015	-0.0003	**-0.0006** **	2.20e-05
		(0.0007)	(0.0002)	(0.0005)	(0.0089)	(0.0028)	(0.0017)	**(0.0003)**	(0.0003)
Cost	Gain	**0.0002** **	**0.0001** **	**-0.0014*** **	0.0001	-2.78e-05	-0.0002	0.0002	-8.05e-05
		(9.31e-05)	**(5.39e-05)**	**(0.0003)**	(0.0007)	(0.0001)	(0.0002)	(0.0001)	(7.23e-05)
	Loss	**0.0002***	**-0.0002*** **	-0.0002	-1.50e-05	6.33e-05	-0.0002	**-3.59e-05*** **	-9.39e-05
		(0.0001)	**(5.13e-05)**	(0.0002)	(5.93e-05)	(0.0002)	(0.0002)	**(1.33e-05)**	(7.25e-05)
Satisfy		**-0.4343***	-0.1826	-0.3713	-0.1292	-0.3371	-0.0331	**-0.5642*** **	**-0.4327*** **

		(0.2234)	(0.1408)	(0.2547)	(0.4572)	(0.4828)	(0.2199)	**(0.1984)**	**(0.1333)**
Constant		**3.2105*** **	**-1.0860*** **	1.3314	0.5225	-1.1811	**-1.1161** **	**3.4104*** **	**0.5294***
		(0.4200)	**(0.3047)**	(0.9920)	(1.1378)	(0.8029)	**(0.4677)**	**(0.3532)**	**(0.2988)**
Observations		202	293	110	26	35	81	326	726
/lnalpha		**-0.5086*** **	**-0.5504*** **	-0.1419	-52.3537	-22.6692	-23.7283	-0.0237	**0.2815*** **
		(0.1274)	**(0.0932)**	(0.1440)	(0)	(0)	(0)	(0.0852)	**(0.0491)**
Log-Likelihood		-557.5909	-1209.3737	-404.6645	-27.0552	-38.7780	-99.2822	-941.9313	-2978.3224

VARIABLES		8AM	9AM	10AM	11AM	12PM	13PM	14PM	15PM
Age		**0.0338*** **	**0.0179*** **	**0.0097***	**0.0613*** **	-9.87e-05	**0.0277*** **	-0.0052	**0.0699*** **
		(0.0042)	**(0.0050)**	**(0.0050)**	**(0.0039)**	(0.0050)	**(0.0042)**	(0.0064)	**(0.0111)**
Gender		-0.1123	**-0.6844*** **	**-0.2157** **	**0.6252*** **	**0.5879*** **	**0.4722*** **	**0.2540***	**0.4024** **
		(0.1121)	**(0.0828)**	**(0.1060)**	**(0.1113)**	**(0.1150)**	**(0.1122)**	**(0.1435)**	**(0.1819)**
WaitTime		**0.0003***	**-0.0005*** **	**-0.0002***	**0.0002***	**-0.0002** **	4.82e-05	**-0.0009*** **	-0.0002
		(0.0001)	**(0.0001)**	**(0.0001)**	**(9.34e-05)**	**(0.0001)**	(0.0001)	**(0.0002)**	(0.0002)
Distance	Gain	-0.0393	0.0367	**-0.0804*** **	**-0.0723***	0.0076	-0.0282	0.0020	0.0033
		(0.0415)	(0.0308)	**(0.0299)**	**(0.0400)**	(0.0389)	(0.0355)	(0.0496)	(0.0391)
	Loss	0.0023	0.0062	-0.0476	**-0.0221***	0.0020	0.0059	-0.0572	-0.0155
		(0.0155)	(0.0085)	(0.0412)	**(0.0114)**	(0.0046)	(0.0187)	(0.0640)	(0.0367)
RideTime	Gain	2.17e-05	**-0.0004** **	-0.0001	-2.38e-05	-0.0004	**-0.0006** **	0.0004	-0.0004
		(0.0003)	**(0.0002)**	(0.0002)	(0.0002)	(0.0003)	**(0.0003)**	(0.0004)	(0.0005)

	Loss	0.0002	**-0.0004****	**-0.0007*****	0.0001	-0.0001	-0.0001	**-0.0008****	-0.0004
		(0.0002)	**(0.0002)**	**(0.0002)**	(8.80e-05)	(0.0001)	(0.0002)	**(0.0003)**	(0.0003)
Cost	Gain	**0.0004*****	**0.0001*****	4.19e-05	4.98e-05	-0.0001	**-0.0002*****	1.70e-06	-0.0002
		(7.72e-05)	**(5.18e-05)**	(6.16e-05)	(7.18e-05)	(0.0001)	**(7.09e-05)**	(0.0001)	(0.0001)
	Loss	**-4.46e-05****	**-0.0001*****	-8.51e-06	-1.03e-05	-1.68e-05	**-7.08e-05***	9.11e-06	-4.61e-06
		(1.79e-05)	**(4.04e-05)**	(3.78e-05)	(1.20e-05)	(1.78e-05)	**(4.27e-05)**	(7.47e-05)	(5.13e-05)
Satisfy		**0.4710*****	**-0.4440*****	-0.0439	**-0.2005***	-0.1492	**-0.1885***	**0.4921*****	0.1413
		(0.1295)	**(0.1125)**	(0.1119)	**(0.1191)**	(0.1239)	**(0.1091)**	**(0.1626)**	(0.1586)
Constant		**1.1128*****	**1.7918*****	**1.1697*****	**-1.7845*****	**0.7004*****	-0.3876	**1.7818*****	**-1.8083*****
		(0.2021)	**(0.2121)**	**(0.2777)**	**(0.1913)**	**(0.2139)**	(0.2378)	**(0.2987)**	**(0.5364)**
Observations		711	659	586	548	531	497	458	211
/lnalpha		**0.2166*****	**-0.3141*****	**-0.5543*****	**-0.7008*****	**-0.6866*****	**-1.2839*****	-0.0413	**-0.7165*****
		(0.0514)	**(0.0656)**	**(0.0834)**	**(0.0913)**	**(0.0951)**	**(0.1357)**	(0.0742)	**(0.1460)**
Log-Likelihood		-2635.2577	-1856.3519	-1319.5716	-1231.7662	-1111.9058	-959.8803	-1190.6452	-456.1036

VARIABLES		16PM	17PM	18PM	19PM	20PM	21PM	22PM	23PM
Age		**0.0315*****	**0.0291*****	-0.0063	-0.0029	**0.0378*****	**0.0376*****	**0.0186*****	**0.0162****
		(0.0077)	**(0.0095)**	(0.0055)	(0.0068)	**(0.0040)**	**(0.0077)**	**(0.0053)**	**(0.0070)**
Gender		**0.5416*****	**1.4912*****	**-0.5678*****	-0.1780	**0.4954*****	**0.5899*****	**0.3364*****	0.2014
		(0.1429)	**(0.1687)**	**(0.0987)**	(0.1272)	**(0.1165)**	**(0.1506)**	**(0.1184)**	(0.1374)
WaitTime		-0.0002	**-0.0008*****	-8.25e-05	3.41e-05	**-0.0005*****	**0.0006*****	**-0.0007*****	**-0.0005****
		(0.0002)	**(0.0002)**	(0.0001)	(0.0001)	**(0.0002)**	**(0.0002)**	**(0.0002)**	**(0.0002)**

Distance	Gain	-0.0066	-0.0035	-0.0119	-0.0590	-0.0315	0.0311	-0.0089	0.0066
		(0.0432)	(0.0541)	(0.0279)	(0.0416)	(0.0241)	(0.0411)	(0.0282)	(0.0228)
	Loss	-0.0105	**-0.1319*****	-0.0296	**-0.0694***	-0.0252	-0.0265	**-0.0470***	-0.0441
		(0.0117)	**(0.0479)**	(0.0254)	**(0.0420)**	(0.0270)	(0.0296)	**(0.0248)**	(0.0475)
RideTime	Gain	-0.0003	-0.0005	-0.0003	**-0.0007***	**0.0006****	**-0.0006***	-0.0002	**-0.0008***
		(0.0003)	(0.0003)	(0.0002)	**(0.0004)**	**(0.0002)**	**(0.0004)**	(0.0003)	**(0.0004)**
	Loss	-0.0003	0.0002	0.0002	-0.0001	0.0002	**-0.0010*****	**0.0005****	**-0.0005***
		(0.0002)	(0.0001)	(0.0002)	(0.0003)	(0.0002)	**(0.0003)**	**(0.0002)**	**(0.00028)**
Cost	Gain	3.03e-05	4.81e-05	7.26e-05	**-0.0002***	**-0.0002*****	-8.64e-05	-9.06e-05	4.85e-05
		(4.96e-05)	(8.93e-05)	(7.05e-05)	**(0.0001)**	**(6.52e-05)**	(6.15e-05)	(6.61e-05)	(6.42e-05)
	Loss	-4.90e-06	**-0.0001***	-6.28e-05	-0.0001	-1.76e-05	**0.0001****	-3.48e-05	-2.46e-05
		(3.12e-05)	**(6.10e-05)**	(4.15e-05)	(8.21e-05)	(4.43e-05)	**(4.61e-05)**	(3.81e-05)	(2.58e-05)
Satisfy		0.1049	**-0.4450****	**-0.6107*****	**-0.4490*****	**-0.4324*****	**0.2911****	-0.1383	**-0.3628***
		(0.1548)	**(0.2035)**	**(0.1409)**	**(0.1658)**	**(0.1489)**	**(0.1405)**	(0.1497)	**(0.2034)**
Constant		**-0.7380****	0.1700	**2.2253*****	**2.1410*****	-0.3091	**-0.8182****	**1.4547*****	**0.7270****
		(0.3741)	(0.3873)	**(0.2244)**	**(0.2825)**	(0.2038)	**(0.3796)**	**(0.2219)**	**(0.2975)**
Observations		212	352	460	352	306	290	385	207
/lnalpha		**-1.5126*****	0.0660	**-0.3579*****	**-0.1678***	**-0.7411*****	**-0.2489****	-0.0012	**-0.8982*****
		(0.2544)	(0.0798)	**(0.0829)**	**(0.0884)**	**(0.1241)**	**(0.1005)**	(0.0773)	**(0.1690)**
Log-Likelihood		-378.3781	-1061.7159	-1155.6832	-920.0649	-723.5587	-682.1803	-1197.9232	-433.6335

*** p<0.01, ** p<0.05, * p<0.1; Standard errors are in parenthesis

Table 34. Summary of significant variables for total usage by times

VARIABLES		0AM	1AM	2AM	3AM	4AM	5AM	6AM	7AM
Age		(-)	(+)	(+)		(+)	(+)	(-)	(+)
Gender		(-)	(-)	(-)				(-)	(+)
WaitTime				(-)					(-)
Distance	Gain		(+)					(-)	
	Loss								
RideTime	Gain	(+)	(-)					(-)	
	Loss							(-)	
Cost	Gain	(+)	(+)	(-)					
	Loss	(+)	(-)					(-)	
Satisfy		(-)						(-)	(-)
Constant		(+)	(-)				(-)	(+)	(+)

VARIABLES		8AM	9AM	10AM	11AM	12PM	13PM	14PM	15PM
Age		(+)	(+)	(+)	(+)		(+)		(+)
Gender			(-)	(-)	(+)	(+)	(+)	(+)	(+)
WaitTime		(+)	(-)	(-)	(+)	(-)		(-)	
Distance	Gain			(-)	(-)				
	Loss				(-)				

RideTime	Gain		(-)				(-)		
	Loss		(-)	(-)				(-)	
Cost	Gain	(+)	(+)				(-)		
	Loss	(-)	(-)				(-)		
Satisfy		(+)	(-)		(-)		(-)	(+)	
Constant		(+)	(+)	(+)	(-)	(+)		(+)	(-)

VARIABLES		16PM	17PM	18PM	19PM	20PM	21PM	22PM	23PM
Age		(+)	(+)			(+)	(+)	(+)	(+)
Gender		(+)	(+)	(-)		(+)	(+)	(+)	
WaitTime			(-)			(-)	(+)	(-)	(-)
Distance	Gain								
	Loss		(-)		(-)			(-)	
RideTime	Gain				(-)	(+)	(-)		(-)
	Loss						(-)	(+)	(-)
Cost	Gain				(-)	(-)			
	Loss		(-)				(+)		
Satisfy			(-)	(-)	(-)	(-)	(+)		(-)
Constant		(-)		(+)	(+)		(-)	(+)	(+)

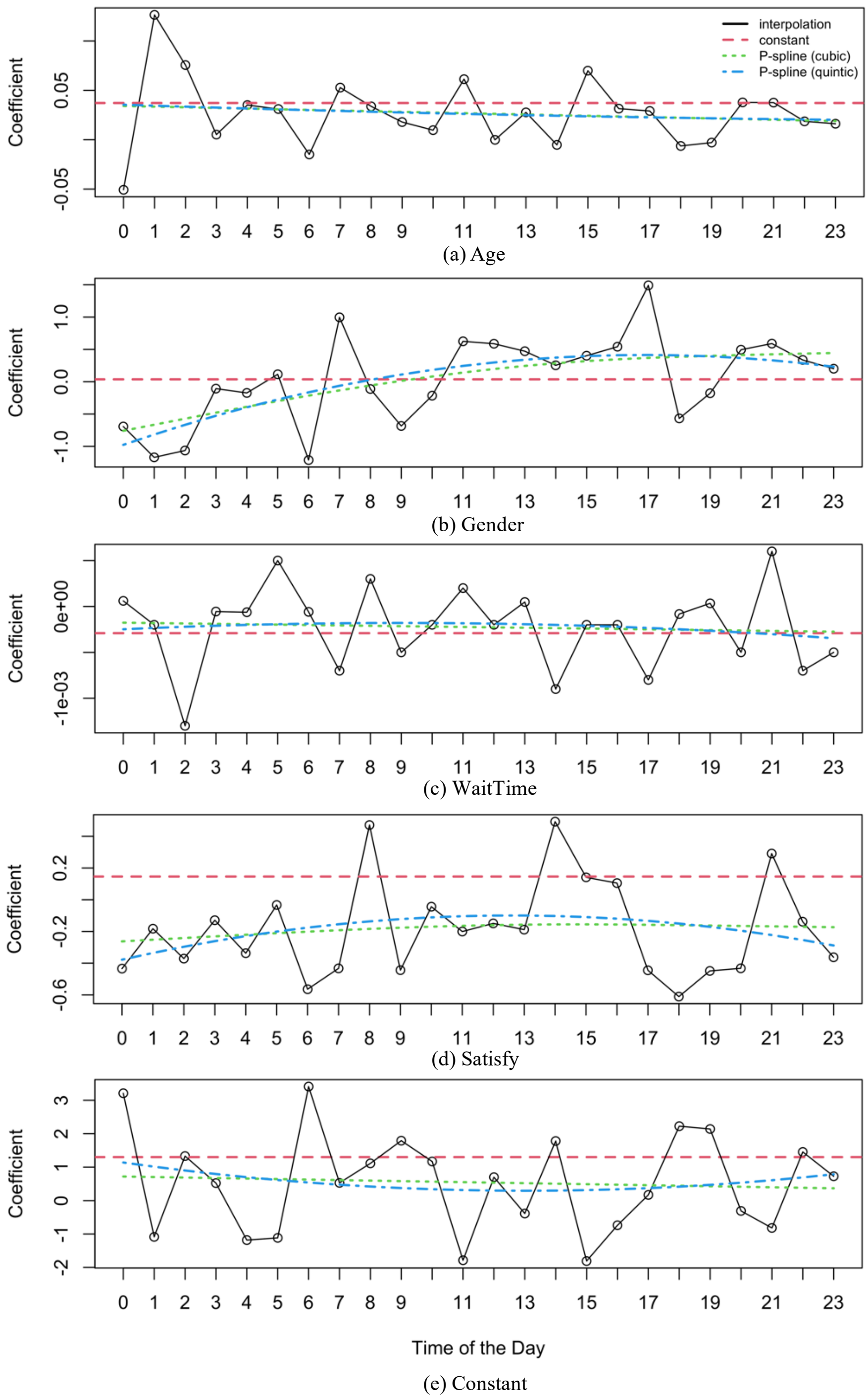


Figure 45. P-spline fits for Total Usage parameters by times of the day

The coefficient trends of gender show that while men have a greater total use of service in mornings until about 8AM-9AM, women's total use of service is greater in the hours following. This suggests that in particular, men use the service more during the late night and early morning hours. On the same line, the level of influence of being men on total use of service consistently decreases with time, and the opposite is true for women. One interesting observation is that lower satisfaction of the service led to the increase in the total number of uses. There exists a possibility that because Macaron Taxi service is still in its infancy having been launched in 2019, users do not yet have high expectations of the service. Therefore, consumers who use Macaron Taxi service to hail rides in substitute of other existing ride-hailing services have higher interest in the newly-developed service and are more willing to leave their comfort zones to invest in the novel service. In those terms, persons who invest their time more in using the novel service may more strictly evaluate the service. Such influence of user satisfaction on total number of service usage is least around noon, and the further the time is from noon, the level of influence increases.

The coverage of the 95% Bayesian confidence interval (CI) for each smoothing was calculated using Eq. (44). Figure 46 shows the estimated functional relationship as well as the simulated coverage of the 95% Bayesian CIs. To add, the summary of fit of the estimated splines is summarized in Table 35 and 36, each showing the approximate significance of parametric and nonparametric effects, respectively.

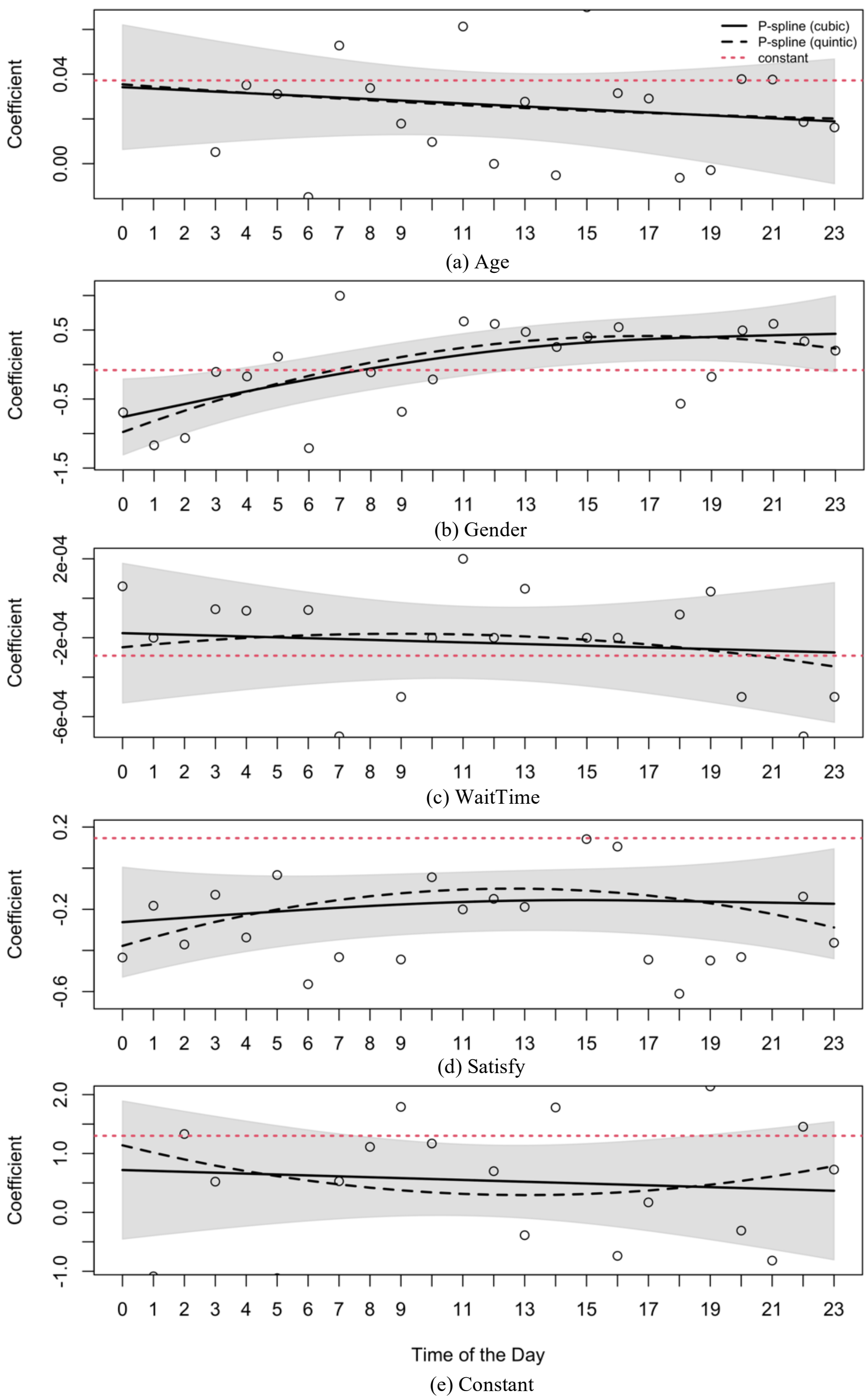


Figure 46. P-spline CIs for Total Usage parameters by times of the day

Putting the results together, it is noticeable that for covariates of Age and $WaitTime$, coefficients that are estimated from the generic model completely falls within the Bayesian CIs of the cubic spline. This indicates that all models sufficiently explain the data, with the nonparametric model showing the long-term variance of time. Likewise, because all nonparametric effects of the covariates are insignificant, this indicates that that the linear effects need to be considered.

For $Gender$, the coefficient that is estimated from the generic model do not fall completely within the Bayesian CIs of the cubic spline. This suggests that the nonparametric model should be preferred in fitting the given coefficients that vary with time. Nonetheless, because the nonparametric effects of the covariate are insignificant, this indicates that that the linear effects need to be considered as the time variances are caused by long-term fluctuations. It should finally be noted that the coefficient of covariate $Satisfy$ was insignificant in the generic model – the constant graph is drawn just for reference, not for interpretation.

Table 35. Parametric effects from P-spline fits of total usage parameters by times

Variables	Parameter	Cubic P-Spline				Quintic P-Spline			
		Estimate	Std. Error	t-value	Pr(>|t|)	Estimate	Std. Error	t-value	Pr(>|t|)
Age	(Intercept)	0.0266	0.0073	3.62	0.0015**	0.0265	0.0075	3.52	0.0021**
	x	-0.0153	0.0244	-0.63	0.5374	-0.0153	0.0249	-0.61	0.5469
	x^2	-	-	-	-	0.0159	0.1858	0.09	0.9326
Gender	(Intercept)	0.0477	0.1155	0.41	0.6841	0.0575	0.1147	0.50	0.6212
	x	1.2038	0.4089	2.94	0.0077**	1.2089	0.3795	3.19	0.0044**
	x^2	-	-	-	-	-5.1592	2.8267	-1.83	0.0822.
WaitTime	(Intercept)	-0.0002	9.325e-05	-2.42	0.0242*	-0.0002	0.0001	-2.33	0.0299*
	x	-0.0001	3.098e-04	-0.32	0.7555	-0.0001	0.0003	-0.31	0.7602
	x^2	-	-	-	-	-0.0009	0.0024	-0.38	0.7074
Satisfy	(Intercept)	-0.1839	0.0632	-2.91	0.0082**	-0.1786	0.0634	-2.82	0.0103*
	x	0.0895	0.2142	0.42	0.6804	0.0896	0.2097	0.43	0.6736
	x^2	-	-	-	-	-1.8541	1.5620	-1.19	0.2485
Constant	(Intercept)	0.5437	0.3087	1.76	0.0921.	0.5247	0.3138	1.67	0.1094
	x	-0.3521	1.0258	-0.34	0.7347	-0.3521	1.0386	-0.34	0.7379
	x^2	-	-	-	-	5.2609	7.7364	0.68	0.5039

*** p<0.001, ** p<0.01, * p<0.05, . p<0.1

Table 36. Nonparam. effects from P-spline fits of total usage parameters by times

		Cubic P-Spline					Quintic P-Spline				
Variables	Parameter	DF	Sum Sq.	Mean Sq.	F-value	Pr(>F)	DF	Sum Sq.	Mean Sq.	F-value	Pr(>F)
Age	$f(x)$	6.56e-07	2.775e-11	4.228e-05	0.03	0.8581	5.34e-09	3.895e-13	7.295e-05	0.05	0.8186
	Residuals	2.20e+01	2.841e-02	1.291e-03	-	-	2.10e+01	2.840e-02	1.352e-03		
Gender	$f(x)$	0.7015	0.5500	0.7840	2.47	0.1318	5.34e-09	3.387e-11	0.0063	0.02	0.8882
	Residuals	21.2985	6.7970	0.3191	-	-	2.10e+01	6.574e+00	0.3130		
WaitTime	$f(x)$	6.56e-07	2.213e-14	3.372e-08	0.16	0.6916	5.34e-09	1.294e-16	2.423e-08	0.11	0.7417
	Residuals	2.20e+01	4.591e-06	2.087e-07	-	-	2.10e+01	4.560e-06	2.171e-07	-	-
Satisfy	$f(x)$	2.2897	0.0368	0.1270	1.32	0.2618	5.34e-09	4.219e-11	0.0079	0.08	0.7765
	Residuals	21.7103	2.0765	0.0957	-	-	2.10e+01	2.007e+00	0.0956	-	-
Constant	$f(x)$	6.56e-07	5.999e-07	0.9142	0.40	0.5338	5.34e-09	6.599e-10	0.1255	0.05	0.8193
	Residuals	2.20e+01	5.032e+01	2.2875	-	-	2.10e+01	4.924e+01	2.3448	-	-

*** $p<0.001$, ** $p<0.01$, * $p<0.05$, . $p<0.1$

Lastly, the service satisfaction ($Satisfy$) of passengers in the Macaron ride-hailing platform service that varies by day was analyzed using the ordered logit (OL). Amongst, Model 1 without the interaction effects of travel distance and speed was used. The estimation results are presented in Table 37 below, and the significant variables for each day are again summarized in Table 38 for clarity. It is notable that for times of 3AM and 4AM, estimation did not converge due to the lack of sample size.

The results reveal that the coefficients for covariates of Age and $Gender$ and were consistently significant on each day of the week, followed by $WaitTime$. Covariates with regards to distance and ride time were almost always insignificant (≤ 6). The estimated paths for coefficients of covariates that were consistently significant and varies by the times of the day were fit using both the cubic and quintic P-spline functions as in Figure 47.

From the figure, it can be seen that the coefficients for age show short-term trend, whereas coefficients for gender and wait time show long-term trends. It is revealed that the older the person, the greater the service satisfaction of ride-hailing platform, and such level of influence is the greatest in late-night/early-morning hours (0AM-6AM) and decreases slowly with time. The quintic spline for age additionally shows a lot of short-term fluctuations, and the level of influence in age increase sharply and reach its the highest at 1AM (most distinct local maxima), and decrease back until 3AM. Although the coefficients fluctuate at other times revealing short-term trends, they are mostly kept at similar levels.

Table 37. Estimation results for satisfaction by times

VARIABLES		0AM	1AM	2AM	3AM	4AM	5AM	6AM	7AM
Age		**-0.0475*** **	**0.2530*** **	**0.1630*** **	-	-	**0.1520*** **	0.0066	**0.0409*** **
		(0.0179)	**(0.0228)**	**(0.0460)**	-	-	**(0.0373)**	(0.0102)	**(0.0076)**
Gender		**-1.2410*** **	**-1.0680*** **	-0.6030	-	-	**1.3030***	**-1.0620*** **	**1.1180*** **
		(0.3130)	**(0.3530)**	(0.6200)	-	-	**(0.7920)**	**(0.2290)**	**(0.1590)**
WaitTime		-0.0004	**-0.0010** **	**-0.0026*** **	-	-	0.0015	-0.0001	**-0.0010*** **
		(0.0003)	**(0.0004)**	**(0.0007)**	-	-	(0.0015)	(0.0005)	**(0.0004)**
Distance	Gain	-0.0711	**0.1550** **	-0.1130	-	-	-0.6010	-0.1240	0.0438
		(0.1140)	**(0.0666)**	(0.1530)	-	-	(0.3750)	(0.0951)	(0.0491)
	Loss	0.1770	0.0074	-0.0103	-	-	-0.8300	0.0504	0.0474
		(0.2120)	(0.0974)	(0.1860)	-	-	(0.8260)	(0.0492)	(0.0409)
RideTime	Gain	**0.0031*** **	**-0.0024*** **	0.0004	-	-	0.0022	**-0.0020*** **	-0.0004
		(0.0008)	**(0.0008)**	(0.0015)	-	-	(0.0025)	**(0.0006)**	(0.0004)
	Loss	-0.0015	0.0002	-0.0002	-	-	-1.6290	**-0.0014***	8.89e-05
		(0.0017)	(0.000563)	(0.0011)	-	-	(161.30)	**(0.0008)**	(0.0005)
Cost	Gain	0.0002	**0.0003*** **	**-0.0030*** **	-	-	**-0.0019***	2.39e-05	**-0.0002***
		(0.0002)	**(0.0001)**	**(0.0008)**	-	-	**(0.0010)**	(0.0002)	**(0.0001)**
	Loss	0.0004	**-0.0005*** **	-0.0004	-	-	**-0.0022***	**-0.0006*** **	-0.0001
		(0.0003)	**(0.0001)**	(0.0004)	-	-	**(0.0012)**	**(0.0002)**	(8.89e-05)
Satisfy $_{t-1}$		**-0.6520***	-0.4400	-0.6480	-	-	0.1480	-0.2830	**-0.7620*** **

	(0.3840)	(0.3180)	(0.5290)	-	-	(0.8750)	(0.3380)	**(0.1960)**
Observations	202	293	110	26	35	81	326	726
/cut1	**-2.424*** **	**6.190*** **	**2.730***	-	-	7.568***	-1.497***	0.425
	(0.790)	**(0.745)**	**(1.444)**	-	-	(2.125)	(0.550)	(0.368)
/cut2	**-1.810** **	**6.618*** **	**2.844** **	-	-	8.766***	-0.942*	0.881**
	(0.784)	**(0.752)**	**(1.446)**	-	-	(2.239)	(0.552)	(0.367)
/cut3	**-1.434***	**6.705*** **	**3.122** **	-	-	9.533***	-0.689	1.055***
	(0.778)	**(0.754)**	**(1.452)**	-	-	(2.355)	(0.552)	(0.367)
/cut4	-0.993	**6.844*** **	**3.525** **	-	-	-	-0.453	1.299***
	(0.770)	**(0.757)**	**(1.458)**	-	-	-	(0.552)	(0.367)
/cut5	-0.500	**6.995*** **	**5.103*** **	-	-	-	-0.222	1.397***
	(0.764)	**(0.760)**	**(1.501)**	-	-	-	(0.551)	(0.367)
Log-Likelihood	-345.4842	-458.1534	-131.5386	-	-	-42.2561	-612.2200	-1900.4877

VARIABLES	8AM	9AM	10AM	11AM	12PM	13PM	14PM	15PM
Age	**0.0415*** **	**0.0287*** **	**0.0336*** **	**0.0904*** **	0.0170	**0.0588*** **	0.0082	**0.1160*** **
	(0.0070)	**(0.0088)**	**(0.0111)**	**(0.0110)**	(0.0104)	**(0.0105)**	(0.0107)	**(0.0238)**
Gender	**-0.2750***	**-0.8800*** **	-0.2720	**1.0780*** **	**1.025*** **	**0.8030*** **	0.1200	**0.8630** **
	(0.1520)	**(0.1650)**	(0.1950)	**(0.2210)**	**(0.2470)**	**(0.2500)**	(0.2060)	**(0.4190)**
WaitTime	0.0002	**-0.0008*** **	**-0.0007** **	0.0003	-0.0002	**-0.0005***	**-0.0011*** **	-0.0002
	(0.0002)	**(0.0002)**	**(0.0003)**	(0.0002)	(0.0002)	**(0.0003)**	**(0.0003)**	(0.0005)

Distance	Gain	-0.0711	0.0750	**-0.1980****	**-0.1680***	0.0062	-0.0198	-0.0414	-0.1580
		(0.0645)	(0.0625)	**(0.0828)**	**(0.0903)**	(0.0890)	(0.0839)	(0.0809)	(0.1330)
	Loss	0.0134	0.0077	-0.0870	-0.0289	0.0007	-0.0362	-0.1560	-0.0492
		(0.0210)	(0.0146)	(0.0840)	(0.0233)	(0.0137)	(0.0602)	(0.1060)	(0.1360)
RideTime	Gain	-0.0004	**-0.0011*****	-0.0004	-0.0001	-0.0009	-0.0010	0.0003	-0.0013
		(0.0004)	**(0.0004)**	(0.0005)	(0.0005)	(0.0007)	(0.0007)	(0.0006)	(0.0010)
	Loss	8.64e-05	**-0.0007***	-0.0009*	0.0002	-0.0002	-0.0003	-0.0006	-9.94e-05
		(0.0003)	**(0.0004)**	(0.0005)	(0.0002)	(0.0004)	(0.0004)	(0.0005)	(0.0009)
Cost	Gain	**0.0005*****	**0.0003*****	0.0002	3.96e-05	-0.0002	**-0.0006*****	7.68e-06	-0.0001
		(0.0001)	**(0.0001)**	(0.0001)	(0.0002)	(0.0002)	**(0.0002)**	(0.0002)	(0.000285)
	Loss	-6.86e-05	**-0.0002****	-1.71e-05	-5.50e-06	-4.32e-05	-0.0002	-4.24e-06	**-0.0004****
		(5.27e-05)	**(6.97e-05)**	(7.41e-05)	(1.88e-05)	(5.95e-05)	(0.0001)	(0.0001)	**(0.0002)**
Satisfy $_{t-1}$		**0.6640*****	-0.1860	0.1520	0.0078	-0.3200	-0.0654	**0.5270****	0.2100
		(0.1870)	(0.1970)	(0.2130)	(0.2230)	(0.2490)	(0.232)	**(0.2370)**	(0.3060)
Observations		711	659	586	548	531	497	458	211
/cut1		**0.546***	-0.580	0.706	**4.339*****	**1.488*****	**2.435*****	-0.0103	**4.794*****
		(0.311)	(0.394)	(0.507)	**(0.503)**	**(0.472)**	**(0.558)**	(0.500)	**(1.153)**
/cut2		**0.989*****	-0.0577	**1.274****	**5.227*****	**2.164*****	**3.249*****	0.528	**5.376*****
		(0.310)	(0.394)	**(0.510)**	**(0.515)**	**(0.478)**	**(0.564)**	(0.501)	**(1.167)**
/cut3		**1.179*****	0.257	**1.763*****	**5.499*****	**2.577*****	**3.589*****	0.801	**5.694*****
		(0.310)	(0.394)	**(0.515)**	**(0.521)**	**(0.483)**	**(0.567)**	(0.502)	**(1.175)**

/cut4		**1.350*** **	0.417	**2.005*** **	**5.675*** **	**2.675*** **	**4.028*** **	**1.165** **	**6.313*** **
		(0.311)	(0.395)	**(0.517)**	**(0.525)**	**(0.484)**	**(0.573)**	**(0.505)**	**(1.190)**
/cut5		**1.562*** **	0.556	**2.116*** **	**5.988*** **	**3.025*** **	**4.705*** **	**1.293** **	**6.574*** **
		(0.313)	(0.395)	**(0.519)**	**(0.532)**	**(0.490)**	**(0.588)**	**(0.505)**	**(1.193)**
Log-Likelihood		-1759.5163	-1478.3713	-969.6323	-883.8076	-791.5655	-713.3038	-806.2947	-298.4029

VARIABLES		16PM	17PM	18PM	19PM	20PM	21PM	22PM	23PM
Age		**0.0420** **	**0.0231** **	-0.00718	2.37e-05	**0.1060*** **	**0.0363*** **	**0.0512*** **	0.0200
		(0.0192)	**(0.0115)**	(0.0105)	(0.0113)	**(0.0120)**	**(0.0125)**	**(0.0093)**	(0.0145)
Gender		**1.0280*** **	**0.9270*** **	**-0.5770*** **	-0.0528	**0.7080*** **	-0.1850	**0.4800** **	0.3640
		(0.3590)	**(0.2290)**	**(0.1890)**	(0.2400)	**(0.2490)**	(0.2530)	**(0.2010)**	(0.2950)
WaitTime		**-0.0012** **	**-0.0013*** **	3.29e-06	9.55e-05	**-0.0007***	0.0003	**-0.0011*** **	**-0.0010***
		(0.0006)	**(0.0003)**	(0.0002)	(0.0003)	**(0.000354)**	(0.0003)	**(0.0003)**	**(0.0005)**
Distance	Gain	-0.1930	-0.02140	-0.0394	-0.0587	**-0.0971***	0.0331	**-0.1690** **	0.0269
		(0.1830)	(0.0642)	(0.0501)	(0.0730)	**(0.0552)**	(0.0785)	**(0.0780)**	(0.0419)
	Loss	-0.1380	**-0.1830***	0.0066	-0.1290	**-0.1550***	-0.0019	**-0.1410** **	-0.0834
		(0.1230)	**(0.0993)**	(0.0419)	(0.0921)	**(0.0871)**	(0.0552)	**(0.0637)**	(0.135)
RideTime	Gain	-0.0003	-0.0002	-0.0007	-0.0006	**0.0012***	-0.0007	-0.0003	**-0.0016***
		(0.0008)	(0.0005)	(0.0005)	(0.0006)	**(0.0006)**	(0.0007)	(0.0007)	**(0.0009)**
	Loss	-0.0007	0.0003	0.0001	-9.66e-05	0.0005	**-0.0013** **	**0.0011*** **	-0.0011
		(0.0006)	(0.0002)	(0.0004)	(0.0005)	(0.0004)	**(0.0005)**	**(0.0004)**	(0.0007)

Cost	Gain	-0.0002	-1.66e-05	0.0002	-0.0002	**-0.0005*** **	-0.0002	-2.04e-06	-5.05e-05
		(0.0003)	(0.0002)	(0.0001)	(0.0002)	**(0.0001)**	(0.0002)	(0.0001)	(0.0001)
	Loss	2.01e-05	**-0.000210****	-0.000120	-0.000143	-0.000108	**0.000179****	5.95e-06	-0.000146
		(0.000111)	**(0.000101)**	(9.81e-05)	(0.000132)	(0.000106)	**(7.54e-05)**	(5.59e-05)	(0.000170)
Satisfy $_{t-1}$		0.1890	-0.4620	**-0.6260*** **	**-0.6450****	-0.2720	0.3330	-0.3310	-0.1240
		(0.3920)	(0.3030)	**(0.2400)**	**(0.2950)**	(0.3110)	(0.2520)	(0.2570)	(0.4410)
Observations		212	352	460	352	306	290	385	207
/cut1		**2.178****	0.499	**-1.210*** **	-0.683	**3.948*** **	**1.341****	**1.086****	0.258
		(0.947)	(0.504)	**(0.456)**	(0.535)	**(0.553)**	**(0.617)**	**(0.425)**	(0.624)
/cut2		**3.192*** **	**0.966***	-0.668	-0.0185	**4.770*** **	**2.147*** **	**1.488*** **	0.834
		(0.960)	**(0.505)**	(0.455)	(0.536)	**(0.567)**	**(0.623)**	**(0.424)**	(0.624)
/cut3		**3.405*** **	**1.333*** **	-0.268	0.132	**5.051*** **	**2.801*** **	**1.682*** **	0.982
		(0.964)	**(0.508)**	(0.456)	(0.537)	**(0.574)**	**(0.633)**	**(0.424)**	(0.626)
/cut4		**3.611*** **	**1.507*** **	0.0688	0.459	**5.378*** **	**3.002*** **	**1.769*** **	**1.814*** **
		(0.969)	**(0.511)**	(0.456)	(0.538)	**(0.583)**	**(0.636)**	**(0.424)**	**(0.642)**
/cut5		**3.900*** **	**1.739*** **	0.245	0.534	**5.603*** **	**3.294*** **	**1.933*** **	**2.012*** **
		(0.978)	**(0.514)**	(0.457)	(0.539)	**(0.590)**	**(0.643)**	**(0.424)**	**(0.649)**
Log-Likelihood		-235.6297	-623.6880	-879.2788	-650.7913	-492.8205	-426.9736	-840.5457	-284.1148

*** p<0.01, ** p<0.05, * p<0.1; Standard errors are in parenthesis

Table 38. Summary of significant variables for satisfaction by times

VARIABLES		0AM	1AM	2AM	3AM	4AM	5AM	6AM	7AM
Age		(-)	(+)	(+)			(+)		(+)
Gender		(-)	(-)				(+)	(-)	(+)
WaitTime			(-)	(-)					(-)
Distance	Gain		(+)						
	Loss								
RideTime	Gain	(+)	(-)					(-)	
	Loss							(-)	
Cost	Gain		(+)	(-)			(-)		(-)
	Loss		(-)				(-)	(-)	
Satisfy $_{t-1}$		(-)							(-)

VARIABLES		8AM	9AM	10AM	11AM	12PM	13PM	14PM	15PM
Age		(+)	(+)	(+)	(+)		(+)		(+)
Gender		(-)	(-)		(+)	(+)	(+)		(+)
WaitTime			(-)	(-)			(-)	(-)	
Distance	Gain			(-)	(-)				
	Loss								
RideTime	Gain		(-)						

	Loss		(-)						
Cost	Gain	(+)	(+)				(-)		
	Loss		(-)						(-)
Satisfy $_{t-1}$		(+)						(+)	

VARIABLES		16PM	17PM	18PM	19PM	20PM	21PM	22PM	23PM
Age		(+)	(+)			(+)	(+)	(+)	
Gender		(+)	(+)	(-)		(+)		(+)	
WaitTime		(-)	(-)			(-)		(-)	(-)
Distance	Gain					(-)		(-)	
	Loss		(-)			(-)		(-)	
RideTime	Gain					(+)			(-)
	Loss						(-)	(-)	
Cost	Gain					(-)			
	Loss		(-)				(+)		
Satisfy $_{t-1}$				(-)	(-)				

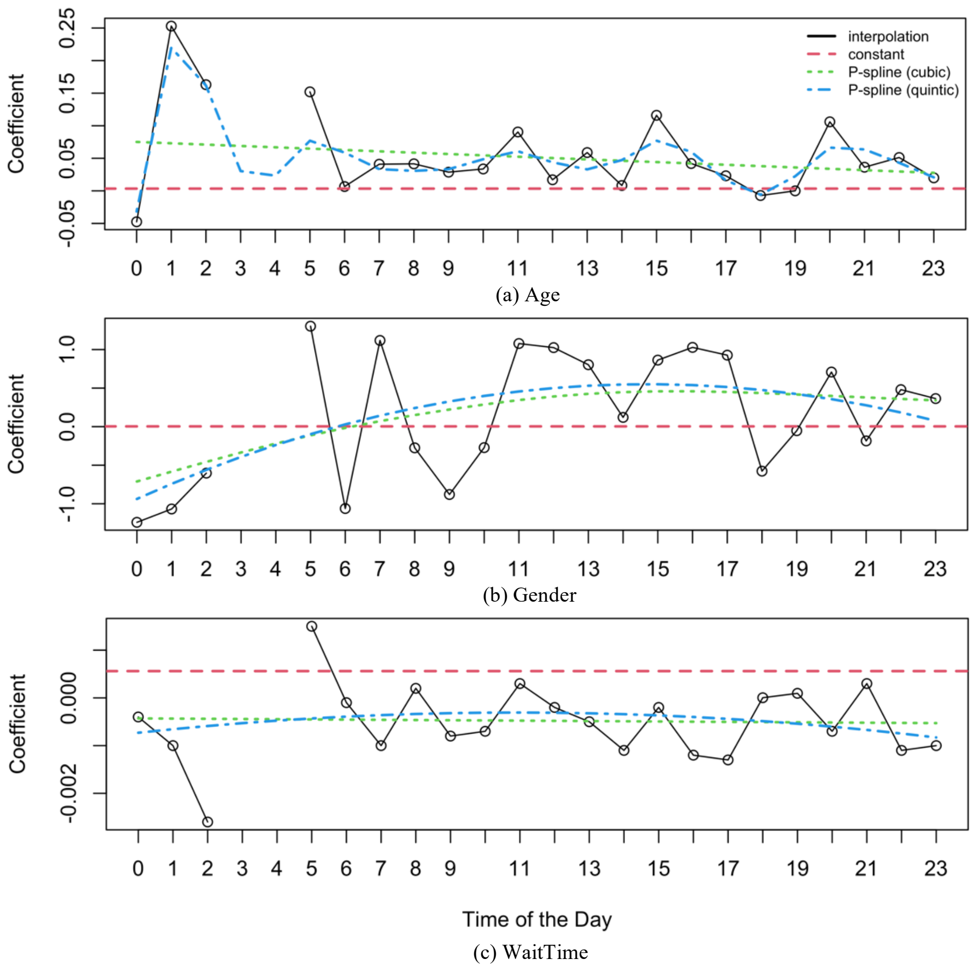
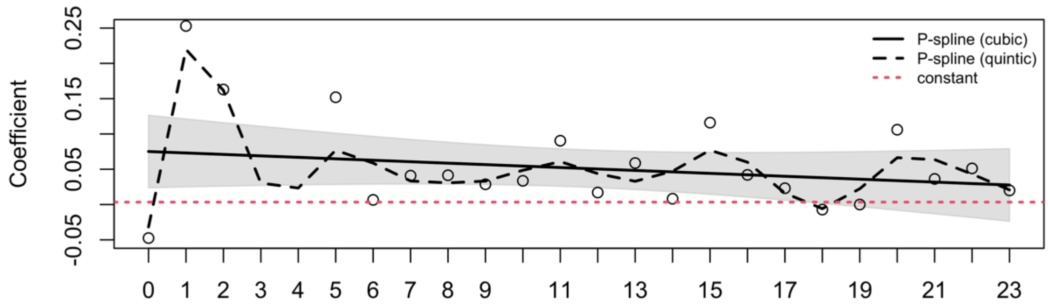


Figure 47. P-spline fits for Satisfaction parameters by times of the day

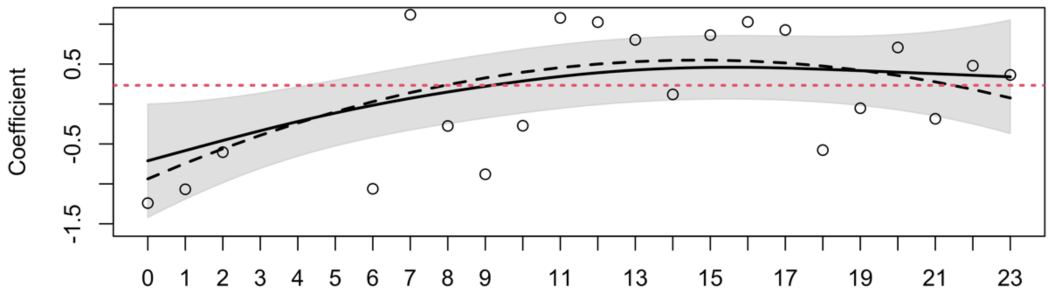
The coefficient trends of gender show that while men have a service satisfaction in late nights from about 0AM to 6AM, women are generally more satisfied in the following hours. This is in accordance with the previous results that men use the service more during the late night and early morning hours. On the same line, the level of influence of being men on service satisfaction consistently decreases with time until about 4PM, and increase back afterwards. The result also indicates that the reduced wait time yield greater service satisfaction, and such tendency is rather consistent with time revealing long-term trends. Such influence is greater in late night and early-morning hours. Overall, it can be observed that the covariate effects on consumer service satisfaction are the greatest from 0AM to 6AM.

The coverage of the 95% Bayesian confidence interval (CI) for each smoothing was calculated using Eq. (44). Figure 48 shows the estimated functional relationship as well as the simulated coverage of the 95% Bayesian CIs. To add, the summary of fit of the estimated splines is summarized in Table 39 and 40, each showing the approximate significance of parametric and nonparametric effects, respectively.

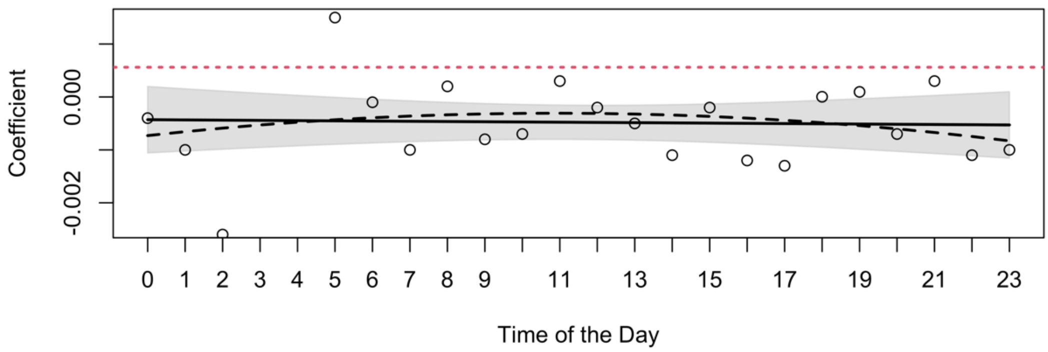
Accordingly, it is noticeable that for all covariates, the coefficient that is estimated from the generic model do not fall completely within the Bayesian CIs of the cubic spline. This suggests that the nonparametric model should be preferred in fitting the given coefficients that vary with time. Nonetheless, because all nonparametric effects of the covariates are insignificant, this indicates that that the linear effects need to be considered. That is because overall, the time variances are generally caused by long-term fluctuations.



(a) Age



(b) Gender



(c) WaitTime

Figure 48. P-spline CIs for Satisfaction parameters by times of the day

Table 39. Parametric effects from P-spline fits of satisfaction parameters by times

		Cubic P-Spline				Quintic P-Spline			
Variables	Parameter	Estimate	Std. Error	t-value	Pr(>|t|)	Estimate	Std. Error	t-value	Pr(>|t|)
Age	(Intercept)	0.0514	0.0135	3.80	0.0010**	0.0556	0.0095	5.86	0.0002***
	x	-0.0475	0.0449	-1.06	0.3018	0.0524	0.0634	0.83	0.4280
	x^2	-	-	-	-	-10.6929	3.7966	-2.82	0.0189*
Gender	(Intercept)	0.1652	0.1421	1.16	0.2581	0.1765	0.1411	1.25	0.2249
	x	1.0516	0.5172	2.03	0.0548.	1.0138	0.4670	2.17	0.0416*
	x^2	-	-	-	-	-7.2880	3.4786	-2.10	0.0485*
WaitTime	(Intercept)	-4.782e-04	0.0002	-2.90	0.0082**	-4.66e-04	0.0002	-2.80	0.0107*
	x	-9.925e-05	0.0005	-0.18	0.8580	-9.93e-05	0.0006	-0.18	0.8586
	x^2	-	-	-	-	-3.77e-03	0.0041	-0.92	0.3680

*** p<0.001, ** p<0.01, * p<0.05, . p<0.1

Table 40. Nonparam. effects from P-spline fits of satisfaction parameters by times

		Cubic P-Spline					Quintic P-Spline				
Variables	Parameter	DF	Sum Sq.	Mean Sq.	F-value	Pr(>F)	DF	Sum Sq.	Mean Sq.	F-value	Pr(>F)
Age	$f(x)$	6.56e-07	8.610e-10	1.312e-03	0.03	0.5897	11.330	0.0636	0.0056	2.69	0.0673.
	Residuals	2.20e+01	9.639e-02	4.381e-03	-	-	9.666	0.0202	0.0021	-	-
Gender	$f(x)$	0.9040	1.3290	1.4701	3.05	0.0955.	5.34e-09	6.943e-10	0.1300	0.27	0.6060
	Residuals	21.0960	10.1790	0.4825	-	-	2.10e+01	9.995e+00	0.4741	-	-
WaitTime	$f(x)$	6.56e-07	3.322e-13	5.062e-07	0.77	0.3900	5.34e-09	2.893e-15	5.417e-07	0.82	0.3745
	Residuals	2.20e+01	1.438e-05	6.535e-07	-	-	2.10e+01	1.382e-05	6.581e-07	-	-

*** p<0.001, ** p<0.01, * p<0.05, . p<0.1

Chapter 6. Conclusion

At the heart of all consumer behaviors are decision-making processes that are influenced by various factors including those that are time-dependent and/or experience-dependent. It is clearly defined in the marketing literature that consumers' intention to repurchase a product or continue to use a service depends primarily on their prior experience of using that product or service, and that continued user satisfaction is considered key to building and retaining a loyal base of long-term customers. However, most existing studies use static utility models to explain consumer behavior in platform services and therefore do not adequately reflect the time-varying effects of continued use of the service. In addition, cross-sectional studies of consumers' continued use of services are unable to capture an accurate view of how customers' expectations and perceptions of the product/service might change over time.

From a managerial point of view, ignoring time-varying effects concerning covariates may leave competitive trends in a product/service category or proper understanding of consumers undetected with the risk of misjudging the nature of consumer behavior. It is further important for service providers to recognize short- or long-term changes of their consumers as a basis for adjusting their marking mix adequately and in due time. This study aimed to fill this gap by employing a dynamic utility model to explain consumer behavior in the platform economy, where services are used repeatedly. Expanding on the expectation-confirmation theory that hypothesizes that a consumer's level of satisfaction with a product/service determines re-purchase intention, and that consumers' continue service use is determined primarily by their satisfaction with prior use of that product or service, this paper incorporated time-varying effects of covariates in explaining consumers' use of platform service using P-splines. Through an empirical study, we examined the time-varying effects of covariates in explaining consumers' use of ride-hailing platforms by first identifying the effect of updating expectations and perceptions with repeated use and then incorporating models based on penalized splines, a semiparametric approach.

In all, this study showed that the dynamics are important in marketing and should be considered in consumer modeling. More specifically, it was shown how the time-varying effect of variables can be applied in analyzing consumer behavior in the widely-used ride-hailing platform service. Through this analysis, it was also clearly shown that implementation of semiparametric model was necessary to identify seasonality of consumer behaviors in platform services. The findings of this paper suggest the necessity to employ models with time-varying parameters and should encourage firms and managers to adopt more flexible models to detect such time-varying effects.

This study also showed how users may make experience-based decisions based on direct or vicarious reinforcement they have received in the past, and how preference for a service is influenced by past usage experiences. Described expectations were readily available from our data, but the experience-based updated expectations and updated perceptions were computed from individual consumers purchase history. While the gap between updated expectation and actual service delivery ("Service Gap") and the gap between updated perception and service delivery ("Perception Gap") were not conceptualized further expanding on the existing GAP model, the results indicated that the 'Service Gap' and 'Perception Gap' sometimes amplifies, and at others, lessens the effect of covariates on consumers' satisfaction of the service. Overall, the study added to the literature by empirically suggesting what marketing implications can be derived through the extended GAP model.

There are several limitations in this study. First, since an empirical analysis was conducted on a very specific service, the managerial implication yielded from the empirical analysis cannot be generalized for all other services. Nonetheless, the study was conducted to show how the time-varying effects of covariates in services with repeated use can be employed. Also, data for only the immediate successful rides were used for the analysis as they are the most used types of service. However, factors such as failure of match after immediate call, or turn-down of user after call can be considered for future research. Also, other types of services provided by the platform, such as reservation of rides, can be observed for further managerial implications.

Bibliography

Ailawadi, K. L., Gedenk, K., & Neslin, S. A. (1999). Heterogeneity and purchase event feedback in choice models: An empirical analysis with implications for model building. *International Journal of Research in Marketing*, *16*(3), 177-198.

Akaike, H. (1974). A new look at the statistical model identification. *IEEE transactions on automatic control*, *19*(6), 716-723.

Alalwan, A. A. (2020). Mobile food ordering apps: An empirical study of the factors affecting customer e-satisfaction and continued intention to reuse. *International Journal of Information Management*, *50*, 28-44.

Alderighi, M., Nava, C. R., Calabrese, M., Christille, J. M., & Salvemini, C. B. (2022). Consumer perception of price fairness and dynamic pricing: Evidence from Booking. com. *Journal of Business Research*, *145*, 769-783.

Allen, D. M. (1974). The relationship between variable selection and data augmentation and a method for prediction. *technometrics*, *16*(1), 125-127.

Anderson, E. W., & Sullivan, M. W. (1993). The antecedents and consequences of customer satisfaction for firms. *Marketing science*, *12*(2), 125-143.

Arthur, W. B. (1989). Competing technologies, increasing returns, and lock-in by historical events. *The economic journal*, *99*(394), 116-131.

Arumugam, V., Ismail, M. R., & Joeharee, M. (2020). A review and conceptual development of the factors influencing consumer intention towards E-hailing service in Malaysia. *International Journal of Innovation, Creativity and*

Change, *11*(11), 224-242.

Ashkrof, P., de Almeida Correia, G. H., Cats, O., & van Arem, B. (2020). Understanding ride-sourcing drivers' behaviour and preferences: Insights from focus groups analysis. *Research in Transportation Business & Management*, *37*, 100516.

Aviv, Y., & Pazgal, A. (2008). Optimal pricing of seasonal products in the presence of forward-looking consumers. *Manufacturing & service operations management*, *10*(3), 339-359.

Bai, J., So, K. C., Tang, C. S., Chen, X., & Wang, H. (2019). Coordinating supply and demand on on-demand service platform with impatient customers. *Manufacturing & Service Operations Management*, *21*(3), 556-570.

Bansback, N., Hole, A. R., Mulhern, B., & Tsuchiya, A. (2014). Testing a discrete choice experiment including duration to value health states for large descriptive systems: addressing design and sampling issues. *Social science & medicine*, *114*, 38-48.

Barney, J. (1991). Firm resources and sustained competitive advantage. *Journal of management*, *17*(1), 99-120.

Bateman, I. J., Day, B. H., Jones, A. P., & Jude, S. (2009). Reducing gain–loss asymmetry: a virtual reality choice experiment valuing land use change. *Journal of environmental economics and management*, *58*(1), 106-118.

Belarmino, A., Raab, C., Tang, J., & Han, W. (2021). Exploring the motivations to use online meal delivery platforms: Before and during quarantine. *International Journal of Hospitality Management*, *96*, 102983.

Bhat, P., Shcherbatyi, I., & Maass, W. (2019). Automated learning of user preferences for selection of high quality 3d designs. *Procedia CIRP*, *84*, 814-819.

Bhattacherjee, A. (2001). Understanding information systems continuance: An expectation-confirmation model. *MIS quarterly*, 351-370.

Bhattacherjee, A., Perols, J., & Sanford, C. (2008). Information technology continuance: A theoretic extension and empirical test. *Journal of Computer Information Systems*, *49*(1), 17-26.

Blattberg, R. C., & Neslin, S. A. (1990). *Sales promotion: Concepts, methods, and strategies*. Prentice Hall.

Bolton, R. N., Lemon, K. N., & Bramlett, M. D. (2006). The effect of service experiences over time on a supplier's retention of business customers. *Management Science*, *52*(12), 1811-1823.

Boulding, W., Kalra, A., Staelin, R., & Zeithaml, V. A. (1993). A dynamic process model of service quality: from expectations to behavioral intentions. *Journal of marketing research*, *30*(1), 7-27.

Brezger, A., & Steiner, W. J. (2008). Monotonic regression based on bayesian p–splines: An application to estimating price response functions from store-level scanner data. *Journal of business & economic statistics*, *26*(1), 90-104.

Brown, S. A., & Venkatesh, V. (2005). A model of adoption of technology in the household: A baseline model test and extension incorporating household life cycle. *Management Information Systems Quarterly*, *29*(3), 11.

Burnap, A., & Hauser, J. (2018). Predicting" Design Gaps" in the Market: Deep Consumer Choice Models under Probabilistic Design Constraints. *arXiv preprint arXiv:1812.11067*.

Carson, R. T., & Groves, T. (2007). Incentive and informational properties of preference questions. *Environmental and resource economics*, *37*(1), 181-210.

Caulfield, B. (2009). Estimating the environmental benefits of ride-sharing: A case study of Dublin. *Transportation Research Part D: Transport and Environment*, *14*(7), 527-531.

Chan, I. C. C., Lam, L. W., Chow, C. W., Fong, L. H. N., & Law, R. (2017). The effect of online reviews on hotel booking intention: The role of reader-reviewer similarity. *International Journal of Hospitality Management*, *66*, 54-65.

Chan, N. D., & Shaheen, S. A. (2012). Ridesharing in North America: Past, present, and future. *Transport reviews*, *32*(1), 93-112.

Chandrasekhar, N., Gupta, S., & Nanda, N. (2019). Food delivery services and customer preference: a comparative analysis. *Journal of Foodservice Business Research*, *22*(4), 375-386.

Chen, G. (2016). A gentle tutorial of recurrent neural network with error backpropagation. *arXiv preprint arXiv:1610.02583*.

Chen, X. M., Zheng, H., Ke, J., & Yang, H. (2020). Dynamic optimization strategies for on-demand ride services platform: Surge pricing, commission rate, and incentives. *Transportation Research Part B: Methodological*, *138*, 23-45.

Chiu, C. M., Chang, C. C., Cheng, H. L., & Fang, Y. H. (2009). Determinants of customer repurchase intention in online shopping. *Online information review*.

Churchill Jr, G. A., & Surprenant, C. (1982). An investigation into the determinants of customer satisfaction. *Journal of marketing research*, *19*(4), 491-504.

Clewlow, R. R., & Mishra, G. S. (2017). Disruptive transportation: The adoption, utilization, and impacts of ride-hailing in the United States.

Craven, P., & Wahba, G. (1978). Smoothing Noisy Data with Spline Functions. Estimating the Correct Degree of Smoothing by the Method of Generalized Cross-Validation. *Numerische Mathematik*, 31, 79.

Cuccia, T., & Rizzo, I. (2011). Tourism seasonality in cultural destinations: Empirical evidence from Sicily. *Tourism management*, *32*(3), 589-595.

Dabholkar, P. A., Shepherd, C. D., & Thorpe, D. I. (2000). A comprehensive framework for service quality: an investigation of critical conceptual and measurement issues through a longitudinal study. *Journal of retailing*, *76*(2), 139-173.

Dai, G., Huang, J., Wambura, S. M., & Sun, H. (2017, May). A balanced assignment mechanism for online taxi recommendation. In *2017 18th IEEE international conference on mobile data management (MDM)* (pp. 102-111). IEEE.

Davis, F. D., Bagozzi, R. P., & Warshaw, P. R. (1989). User acceptance of computer technology: A comparison of two theoretical models. *Management science*, *35*(8), 982-1003.

De Boor, C. (2001). Calculation of the smoothing spline with weighted roughness measure.

Mathematical Models and Methods in Applied Sciences, *11*(01), 33-41.

DellaVigna, S. (2009). Psychology and economics: Evidence from the field. *Journal of Economic literature*, *47*(2), 315-72.

Dhar, R. (1997). Consumer preference for a no-choice option. *Journal of consumer research*, *24*(2), 215-231.

Dickinger, A., & Mazanec, J. A. (2008, January). Consumers' preferred criteria for hotel online booking. In *ENTER* (pp. 244-254).

Duro, J. A., & Turrion-Prats, J. (2019). Tourism seasonality worldwide. *Tourism Management Perspectives*, *31*, 38-53.

Eilers, P. H., Marx, B. D., & Durbán, M. (2015). Twenty years of P-splines. *Sort*, *39*(2), 149-186.

Farrell, J., & Klemperer, P. (2007). Coordination and lock-in: Competition with switching costs and network effects. *Handbook of industrial organization*, *3*, 1967-2072.

Fazio, R. H., & Zanna, M. P. (1981). Direct experience and attitude-behavior consistency. In *Advances in experimental social psychology* (Vol. 14, pp. 161-202). Academic Press.

Ferguson, E. (1997). The rise and fall of the American carpool: 1970–1990. *Transportation*, *24*(4), 349-376.

Fernández-Morales, A., Cisneros-Martínez, J. D., & McCabe, S. (2016). Seasonal concentration of tourism demand: Decomposition analysis and marketing implications. *Tourism Management*, *56*, 172-190.

Flynn, T. N. (2010). Using conjoint analysis and choice experiments to estimate QALY values. *Pharmacoeconomics*, 28(9), 711-722.

Frei, C., Hyland, M., & Mahmassani, H. S. (2017). Flexing service schedules: Assessing the potential for demand-adaptive hybrid transit via a stated preference approach. *Transportation Research Part C: Emerging Technologies*, *76*, 71-89.

Fu, X. M., Zhang, J. H., & Chan, F. T. (2018). Determinants of loyalty to public transit: A model integrating Satisfaction-Loyalty Theory and Expectation-Confirmation Theory. *Transportation Research Part A: Policy and Practice*, *113*, 476-490.

Ganesh, J., Arnold, M. J., & Reynolds, K. E. (2000). Understanding the customer base of service providers: an examination of the differences between switchers and stayers. *Journal of marketing*, *64*(3), 65-87.

Gao, J., Ranjbari, A., & MacKenzie, D. (2019). Would being driven by others affect the value of travel time? Ridehailing as an analogy for automated vehicles. *Transportation*, *46*(6), 2103-2116.

Guadagni, P. M., & Little, J. D. (1983). A logit model of brand choice calibrated on scanner data. *Marketing science*, *2*(3), 203-238.

Gupta, A., Dhiman, N., Yousaf, A., & Arora, N. (2021). Social comparison and continuance intention of smart fitness wearables: An extended expectation confirmation theory perspective. *Behaviour & Information Technology*, *40*(13), 1341-1354.

Haaijer, R., Kamakura, W. A., & Wedel, M. (2001). The 'no-choice' alternative in conjoint choice experiments. *International Journal of Market Research*, *43*.

Habib, A., Irfan, M., & Shahzad, M. (2022). Modeling the enablers of online consumer engagement and platform preference in online food delivery platforms during COVID-19. *Future Business Journal*, *8*(1), 1-18.

Hardie, B. G., Johnson, E. J., & Fader, P. S. (1993). Modeling loss aversion and reference dependence effects on brand choice. *Marketing science*, *12*(4), 378-394.

Härdle, W. (1990). *Applied nonparametric regression* (No. 19). Cambridge university press.

Hastie, T., & Tibshirani, R. (1993). Varying-coefficient models. *Journal of the Royal Statistical Society: Series B (Methodological)*, *55*(4), 757-779.

Haupt, H., Kagerer, K., & Steiner, W. J. (2014). Smooth Quantile-Based Modeling Of Brand Sales, Price And Promotional Effects From Retail Scanner Panels. *Journal of Applied Econometrics*, *29*(6), 1007-1028.

Helson, H. (1964). Adaptation-level theory: An experimental and systematic approach to behavior.

Hess, S., Fowler, M., Adler, T., & Bahreinian, A. (2012). A joint model for vehicle type and fuel type choice: evidence from a cross-nested logit study. *Transportation*, *39*(3), 593-625.

Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. *Neural computation*, *9*(8), 1735-1780.

Hsiao, C. H., Chang, J. J., & Tang, K. Y. (2016). Exploring the influential factors in continuance usage of mobile social Apps: Satisfaction, habit, and customer value perspectives. *Telematics and Informatics*, *33*(2), 342-355.

Hu, M., & Zhou, Y. (2022). Dynamic type matching. *Manufacturing & Service Operations Management*, *24*(1), 125-142.

Hylleberg, S. (1992). *Modelling seasonality*. Oxford University Press.

Ibrahim, R. (2019). On Queues with a Random Capacity: Some Theory, and an Application. In *Sharing Economy* (pp. 279-316). Springer, Cham.

Jonker, M. F., Donkers, B., de Bekker-Grob, E. W., & Stolk, E. A. (2018). Advocating a paradigm shift in health-state valuations: the estimation of time-preference corrected QALY tariffs. *Value in Health*, *21*(8), 993-1001.

Kahneman, D., & Tversky, A. (1979). On the interpretation of intuitive probability: A reply to Jonathan Cohen.

Kahneman, D., & Tversky, A. (2013). Prospect theory: An analysis of decision under risk. *In Handbook of the fundamentals of financial decision making: Part I* (pp. 99-127).

Kakao Mobility Report. (2019). 2019 Kakao Mobility Report. Retrieved from: https://www.kakaomobility.com/contents/5#mobility-report

Kalwani, M. U., & Narayandas, N. (1995). Long-term manufacturer-supplier relationships: do they pay off for supplier firms?. *Journal of marketing*, *59*(1), 1-16.

Kalyanaram, G., & Little, J. D. (1994). An empirical analysis of latitude of price acceptance in consumer package goods. *Journal of consumer research*, *21*(3), 408-418.

Kalyanaram, G., & Winer, R. S. (1995). Empirical generalizations from reference price research. *Marketing science*, *14*(3_supplement), G161-G169.

Keane, M. P., & Wolpin, K. I. (2009). Empirical applications of discrete choice dynamic programming models. *Review of Economic Dynamics*, *12*(1), 1-22.

Kelly, K. L. (2007). Casual carpooling-enhanced. *Journal of Public Transportation*, *10*(4), 6.

Khotimah, D. A. K., & Sarno, R. (2019). Sentiment analysis of hotel aspect using probabilistic latent semantic analysis, word embedding and LSTM. *International Journal of Intelligent Engineering and Systems*, *12*(4), 275-290.

Kim, J., Seung, H., Lee, J., & Ahn, J. (2020). Asymmetric preference and loss aversion for electric vehicles: The reference-dependent choice model capturing different preference directions. *Energy Economics*, *86*, 104666.

Kim, S. (2021, Sept 14). "93% of taxi drivers nationwide have signed up for Kakao T"... de facto monopoly. *Maeil Business Newspaper*. https://www.mk.co.kr/news/it/view/2021/09/885039/.

Klemperer, P. (1987). Markets with consumer switching costs. *The quarterly journal of economics*, *102*(2), 375-394.

Klemperer, P. (1989). Price wars caused by switching costs. *The Review of Economic Studies*, *56*(3), 405-420.

Kneib, T., Baumgartner, B., & Steiner, W. J. (2007). Semiparametric multinomial logit models for analysing consumer choice behaviour. *AStA Advances in Statistical Analysis*, *91*(3), 225-244.

Lee, M. C. (2010). Explaining and predicting users' continuance intention toward e-learning: An extension of the expectation–confirmation model. *Computers &*

Education, *54*(2), 506-516.

Li, K. C. (1987). Asymptotic optimality for Cp, CL, cross-validation and generalized cross-validation: discrete index set. *The Annals of Statistics*, 958-975.

Lu, H. P., & Su, P. Y. J. (2009). Factors affecting purchase intention on mobile shopping web sites. *Internet Research*.

Luo, Y., & Xu, X. (2021). Comparative study of deep learning models for analyzing online restaurant reviews in the era of the COVID-19 pandemic. *International Journal of Hospitality Management*, *94*, 102849.

Martín, J. M. M., Aguilera, J. D. D. J., & Moreno, V. M. (2014). Impacts of seasonality on environmental sustainability in the tourism sector based on destination type: An application to Spain's Andalusia region. *Tourism Economics*, *20*(1), 123-142.

Martínez, J. M. G., Martín, J. M. M., & Rey, M. D. S. O. (2020). An analysis of the changes in the seasonal patterns of tourist behavior during a process of economic recovery. *Technological Forecasting and Social Change*, *161*, 120280.

Masiero, L., Viglia, G., & Nieto-Garcia, M. (2020). Strategic consumer behavior in online hotel booking. *Annals of Tourism Research*, *83*, 102947.

Marx, B. D., & Eilers, P. H. (1999). Generalized linear regression on sampled signals and curves: a P-spline approach. *Technometrics*, *41*(1), 1-13.

Morgan Stanley. (2016, Jun). Shared Mobility on the Road of the Future. *Global Investment Implications of Auto 2.0*. https://www.morganstanley.com/ideas/car-of-future-is-autonomous-electric-shared-mobility

Mulhern, B., Bansback, N., Hole, A. R., & Tsuchiya, A. (2017). Using discrete choice experiments with duration to model EQ-5D-5L health state preferences: testing experimental design strategies. *Medical Decision Making*, *37*(3), 285-297.

Nadal, J. R., Font, A. R., & Rossello, A. S. (2004). The economic determinants of seasonal patterns. *Annals of Tourism Research*, *31*(3), 697-711.

Nguyen-Phuoc, D. Q., Su, D. N., Tran, P. T. K., Le, D. T. T., & Johnson, L. W. (2020). Factors influencing customer's loyalty towards ride-hailing taxi services–A case study of Vietnam. *Transportation Research Part A: Policy and Practice*, *134*, 96-112.

Norman, R., Cronin, P., & Viney, R. (2013). A pilot discrete choice experiment to explore preferences for EQ-5D-5L health states. *Applied health economics and health policy*, *11*(3), 287-298.

Nychka, D. (1988). Bayesian confidence intervals for smoothing splines. *Journal of the American Statistical Association*, *83*(404), 1134-1143.

Oh, S., Ji, H., Kim, J., Park, E., & del Pobil, A. P. (2022). Deep learning model based on expectation-confirmation theory to predict customer satisfaction in hospitality service. *Information Technology & Tourism*, *24*(1), 109-126.

Oliver, R. L. (1980). A cognitive model of the antecedents and consequences of satisfaction decisions. *Journal of marketing research*, *17*(4), 460-469.

Oliver, R. L. (1993). Cognitive, affective, and attribute bases of the satisfaction response. *Journal of consumer research*, *20*(3), 418-430.

Oliver, R. L. (1999). Whence consumer loyalty?. *Journal of marketing*, *63*, 33-44.

Oliver, R. L., & Linda, G. (1981). Effect of satisfaction and its antecedents on consumer preference and intention. *ACR North American Advances*.

Parasuraman, A., Zeithaml, V. A., & Berry, L. L. (1985). A conceptual model of service quality and its implications for future research. *Journal of marketing*, *49*(4), 41-50.

Park, K., Ha, J., & Park, J. Y. (2017). An experimental investigation on the determinants of online hotel booking intention. *Journal of Hospitality Marketing & Management*, *26*(6), 627-643.

Parker, G., Smith, L., & Van Alstyne, M. (2010). Growth and innovation in platform ecosystems. *Tulane University, University of Michigan, and Boston University*.

Parker, G. G., Van Alstyne, M. W., & Choudary, S. P. (2016). *Platform revolution: How networked markets are transforming the economy and how to make them work for you*. WW Norton & Company.

Patterson, H. D., & Thompson, R. (1971). Recovery of inter-block information when block sizes are unequal. *Biometrika*, *58*(3), 545-554.

Patterson, P. G., & Spreng, R. A. (1997). Modelling the relationship between perceived value, satisfaction and repurchase intentions in a business-to-business, services context: an empirical examination. *International Journal of service Industry management*.

Ramsay, J. O., & Silverman, B. W. (2002). *Applied functional data analysis: methods and case studies* (Vol. 77, pp. 115-130). New York: Springer.

Rahi, S., & Ghani, M. A. (2019). Integration of expectation confirmation theory and self-determination theory in internet banking continuance intention. *Journal of Science and Technology Policy Management*.

Ray, A., Dhir, A., Bala, P. K., & Kaur, P. (2019). Why do people use food delivery apps (FDA)? A uses and gratification theory perspective. *Journal of Retailing and Consumer Services*, *51*, 221-230.

Reichheld, F. F., & Sasser, W. E. (1990). Zero Defections: Quality comes to services. *Harvard business review*, *68*(5), 105-111.

Roh, M., & Park, K. (2019). Adoption of O2O food delivery services in South Korea: The moderating role of moral obligation in meal preparation. *International Journal of Information Management*, *47*, 262-273.

Ruppert, D., Wand, M. P., & Carroll, R. J. (2003). Semiparametric Regression.,(Cambridge University Press: Cambridge, UK.).

Rust, J. (1987). Optimal replacement of GMC bus engines: An empirical model of Harold Zurcher. *Econometrica: Journal of the Econometric Society*, 999-1033.

Rust, R. T., & Oliver, R. L. (1994). Service quality: insights and managerial implications from the frontier. *Service quality: New directions in theory and practice*, *7*(12), 1-19.

Schwarz, G. (1978). Estimating the dimension of a model. *The annals of statistics*, 461-464.

Shapiro, C., Varian, H. R., & Carl, S. (1998). Information rules: *A strategic guide to the* *network economy*. Harvard Business Press.

Soysal, G. P., & Krishnamurthi, L. (2012). Demand dynamics in the seasonal goods industry: An empirical analysis. *Marketing Science*, *31*(2), 293-316.

Spreng, R. A., MacKenzie, S. B., & Olshavsky, R. W. (1996). A reexamination of the determinants of consumer satisfaction. *Journal of marketing*, *60*(3), 15-32.

Statistica. (2022, Jun). Number of rides Uber gave worldwide from Q2 2017 to Q4 2020. Statistica Research Department. Retrieved from: https://www.statista.com/statistics/946298/uber-ridership-worldwide/.

Steiner, W. J., Brezger, A., & Belitz, C. (2007). Flexible estimation of price response functions using retail scanner data. *Journal of retailing and consumer services*, *14*(6), 383-393.

Steiner, W. J., Siems, F. U., Weber, A., & Guhl, D. (2014). How customer satisfaction with respect to price and quality affects customer retention: an integrated approach considering nonlinear effects. *Journal of Business Economics*, *84*(6), 879-912.

Stone, M. (1974). Cross-validatory choice and assessment of statistical predictions. *Journal of the royal statistical society: Series B (Methodological)*, *36*(2), 111-133.

Suhartanto, D., Helmi Ali, M., Tan, K. H., Sjahroeddin, F., & Kusdibyo, L. (2019). Loyalty toward online food delivery service: the role of e-service quality and food quality. *Journal of foodservice business research*, *22*(1), 81-97.

Sun, L., Teunter, R. H., Hua, G., & Wu, T. (2020). Taxi-hailing platforms: Inform or Assign drivers?. *Transportation Research Part B: Methodological*, *142*, 197-212.

Tam, C., Santos, D., & Oliveira, T. (2020). Exploring the influential factors of continuance intention to use mobile Apps: Extending the expectation confirmation model. *Information Systems Frontiers*, *22*(1), 243-257.

Thong, J. Y., Hong, S. J., & Tam, K. Y. (2006). The effects of post-adoption beliefs on the expectation-confirmation model for information technology continuance. *International Journal of human-computer studies*, *64*(9), 799-810.

Tversky, A., & Kahneman, D. (1991). Loss aversion in riskless choice: A reference-dependent model. *The quarterly journal of economics*, *106*(4), 1039-1061.

Tse, D. K., & Wilton, P. C. (1988). Models of consumer satisfaction formation: An extension. *Journal of marketing research*, *25*(2), 204-212.

Valvi, A. C., & West, D. C. (2013). E-loyalty is not all about trust, price also matters: extending expectation-confirmation theory in bookselling websites. *Journal of Electronic Commerce Research*, *14*(1), 99.

Van der Heijden, H. (2004). User acceptance of hedonic information systems. *MIS quarterly*, 695-704.

Vaswani, K. (2021, May 3). Grab: How an Uber killer became a powerful Asian super app. *BBC.com*. https://www.bbc.com/news/business-56967633.

Venkatesh, V., Thong, J. Y., & Xu, X. (2012). Consumer acceptance and use of information technology: extending the unified theory of acceptance and use of technology. *MIS quarterly*, 157-178.

Wahba, G. (1983). Bayesian “confidence intervals” for the cross-validated smoothing spline. *Journal of the Royal Statistical Society: Series B (Methodological)*, *45*(1), 133-150.

Wahba, G. (1985). A comparison of GCV and GML for choosing the smoothing parameter in the generalized spline smoothing problem. *The annals of statistics*, 1378-1402.

Wahba, G., & Wold, S. (1975). A completely automatic french curve: fitting spline functions by cross validation. *Communications in Statistics-Theory and Methods*, *4*(1), 1-17.

Wang, Y. (1998). Smoothing spline models with correlated random errors. *Journal of the American Statistical Association*, *93*(441), 341-348.

Wang, Y., Wu, C., & Zhu, T. (2019). Mobile hailing technology and taxi driving behaviors. *Marketing Science*, *38*(5), 734-755.

Wei, X., Nan, G., Dou, R., & Li, M. (2020). Optimal business model for the monopolistic ride-hailing platform: Pooling, premier, or hybrid?. *Knowledge-Based Systems*, *204*, 106093.

Westbrook, R. A., & Reilly, M. D. (1983). Value-percept disparity: an alternative to the disconfirmation of expectations theory of consumer satisfaction. *ACR North American Advances*.

Wood, S. N. (2017). P-splines with derivative based penalties and tensor product smoothing of unevenly distributed data. *Statistics and Computing*, *27*(4), 985-989.

Xi, M., Luo, Z., Wang, N., & Yin, J. (2019). A latent feelings-aware rnn model for user churn prediction with behavioral data. *arXiv preprint arXiv:1911.02224*.

Yan, C., Zhu, H., Korolko, N., & Woodard, D. (2020). Dynamic pricing and matching in ride-hailing platforms. *Naval Research Logistics (NRL)*, *67*(8), 705-724.

Yu, B. (2021). How consumer opinions are affected by marketers: an empirical examination by deep learning approach. *Journal of Research in Interactive Marketing*.

Appendix: Stream-of-Time Effects (Days of the Week)

In the Appendix, the stream-of-time effects are additionally observed by the days of the week to observe the seasonality effect in explaining consumers' use of ride-hailing platforms. Again, the data pooled from Macaron Taxi totaling 8,564 successful immediate rides were categorized by the days of the week to analyze stream-of-time effects by the days of the week. The number of successful rides as categorized by days are presented in Figure Appendix 1.

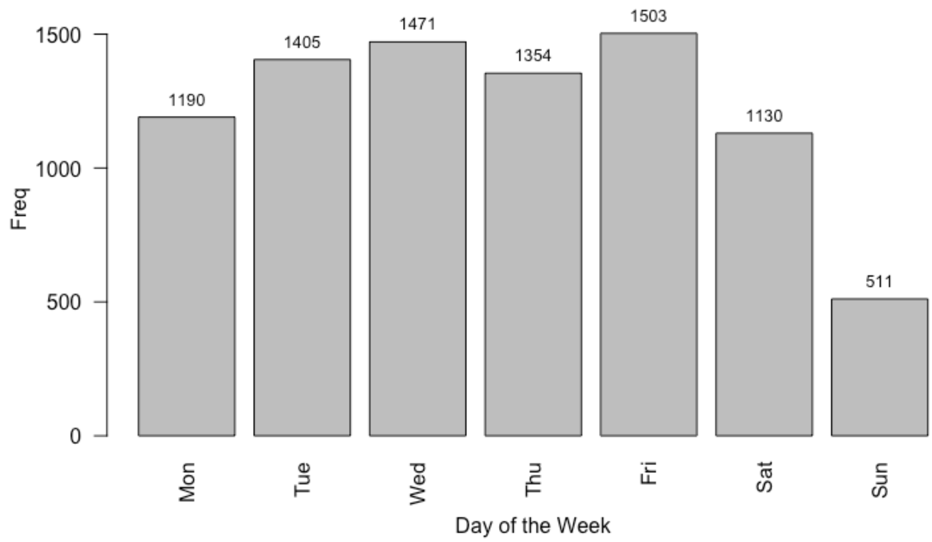


Figure Appendix 1. Service usage frequency by days of the week

Alike the generic model, to investigate the usage interval and the total number of usage of passengers in the Macaron ride-hailing platform service, negative binomial regression was used. The same empirical models used to analyze the stream-of-time effects of by the times of the day was used for the analysis of stream-of-time effects by the days of the week. Our interest in using Model 1 is to identify the influence of gains and losses of specified covariates on the interval of consumers' use of service as well as their total number of usages. For usage interval analysis ($IntvlUse$), consumers with only a single use of service were excluded. The estimation results for usage interval are presented in Table Appendix 1, and the significant variables for each day are again summarized in Table Appendix 2.

The results reveal that the coefficients for covariates of $Gender$ and $Distance\_Gain_{t-1}$ were consistently significant on most days of the week, and the constant was significant on all days. The variables with regards to cost were insignificant at all times. Coefficients for covariates that were consistently significant and varies by day of the week are fit using both the cubic and quintic P-spline functions as shown in Figure Appendix 2. It depicts estimated coefficient with calendar time τ in days on the horizontal axis and estimated coefficients for covariate effects and constant term on the vertical axis. The solid line represents full interpolation of estimated coefficients, and the constant coefficient estimated in the generic model is also depicted. Cubic and quintic splines are those with two lowest degrees that allows separate control on the two end points and two end derivatives, while being the lowest degree that allows reflection points. While cubic splines are most popular as it is the lowest degree meeting, we also incorporate quintic splines to observe a smoother curve at the expense of additional derivatives.

Table Appendix 1. Estimation results for usage interval by days

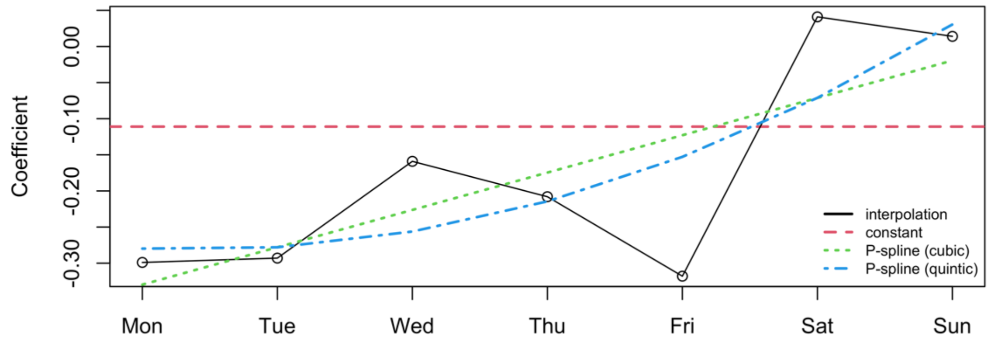
VARIABLES		Monday	Tuesday	Wednesday	Thursday	Friday	Saturday	Sunday
Age		**-0.00705***	**-0.00825***	-0.000851	0.00288	-0.00519	0.00241	-0.00142
		(0.00361)	**(0.00439)**	(0.00480)	(0.00492)	(0.00402)	(0.00608)	(0.00879)
Gender		**-0.299*****	**-0.293*****	**-0.159***	**-0.208****	**-0.318*****	0.0410	0.0140
		(0.0973)	**(0.0916)**	**(0.0933)**	**(0.0893)**	**(0.0919)**	(0.115)	(0.187)
WaitTime $_{t-1}$		0.000186	-3.82e-05	-0.000115	-9.60e-05	**0.000226****	-6.97e-05	-6.30e-05
		(0.000115)	(0.000135)	(0.000135)	(0.000143)	**(0.000108)**	(0.000131)	(0.000315)
Distance	Gain $_{t-1}$	**-0.0631****	**-0.0490****	**-0.0501***	**-0.0699****	0.0151	-0.0324	0.00412
		(0.0263)	**(0.0235)**	**(0.0274)**	**(0.0288)**	(0.0235)	(0.0355)	(0.0526)
	Loss $_{t-1}$	0.0102	**0.0334***	0.0207	0.0171	0.0107	**0.0330***	0.0234
		(0.0312)	**(0.0177)**	(0.0194)	(0.0242)	(0.0220)	**(0.0192)**	(0.0844)
RideTime	Gain $_{t-1}$	0.000163	0.000247	**0.000380***	0.000205	-0.000314	1.30e-05	0.000348
		(0.000286)	(0.000256)	**(0.000230)**	(0.000224)	(0.000234)	(0.000300)	(0.000483)
	Loss $_{t-1}$	9.61e-06	0.000248	0.000208	0.000119	-9.06e-05	2.40e-05	**0.000954***
		(0.000161)	(0.000211)	(0.000171)	(0.000218)	(0.000136)	(0.000265)	**(0.000564)**
Cost	Gain $_{t-1}$	8.18e-05	-5.11e-05	-5.41e-05	6.02e-05	3.03e-05	-1.90e-05	-2.23e-06
		(6.26e-05)	(6.55e-05)	(6.17e-05)	(5.64e-05)	(6.00e-05)	(4.33e-05)	(0.000148)
	Loss $_{t-1}$	-1.37e-05	-3.01e-05	-1.65e-05	-3.36e-05	1.93e-05	4.73e-05	-1.25e-05
		(1.18e-05)	(4.97e-05)	(1.32e-05)	(4.10e-05)	(4.00e-05)	(5.88e-05)	(8.01e-05)
Satisfy $_{t-1}$		0.214	0.326	0.0631	0.0231	0.202	**0.823*****	-0.384

	(0.223)	(0.234)	(0.230)	(0.247)	(0.213)	**(0.305)**	(0.428)
Constant	**2.904*** **	**3.137*** **	**2.850*** **	**2.799*** **	**3.050*** **	**2.268*** **	**2.963*** **
	(0.231)	**(0.262)**	**(0.270)**	**(0.255)**	**(0.244)**	**(0.321)**	**(0.433)**
Observations	621	741	736	663	741	441	186
/lnalpha	-0.0945	0.0635	**0.136*** **	0.0382	**0.131*** **	**0.203*** **	0.110
	(0.0575)	(0.0504)	**(0.0504)**	(0.0535)	**(0.0500)**	**(0.0644)**	(0.100)
Log-Likelihood	-2243.2735	-2816.0188	-2766.2696	-2516.7999	-2835.4810	-1708.9624	-712.9231

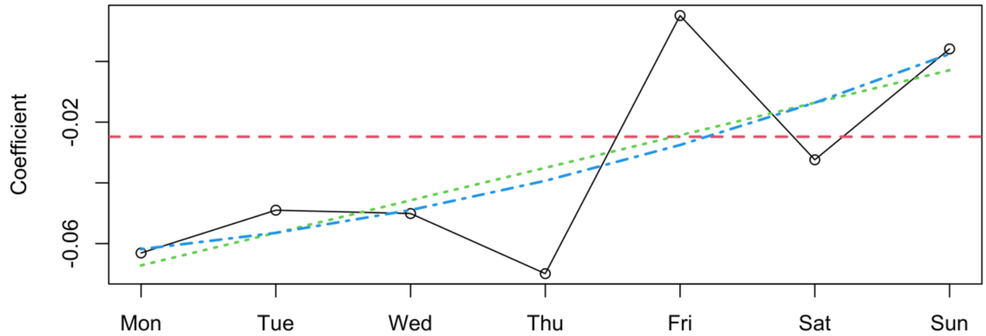
*** p<0.01, ** p<0.05, * p<0.1; Standard errors are in parenthesis

Table Appendix 2. Summary of significant variables for usage interval by days

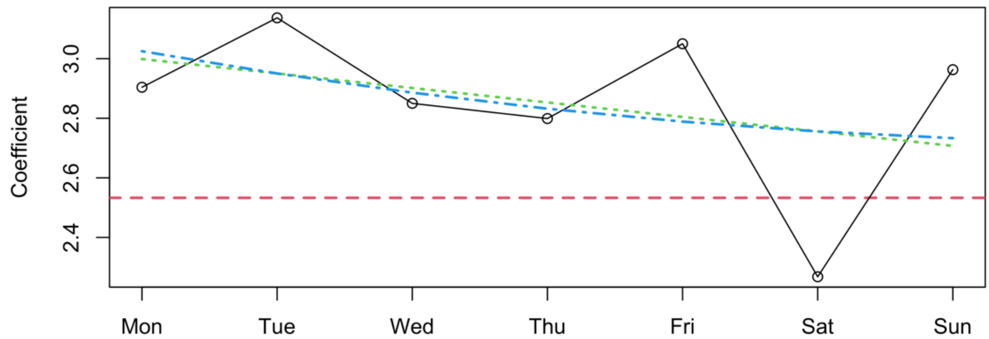
VARIABLES		Monday	Tuesday	Wednesday	Thursday	Friday	Saturday	Sunday
Age		**(-)**	**(-)**					
Gender		**(-)**	**(-)**	**(-)**	**(-)**	**(-)**		
WaitTime $_{t-1}$						**(+)**		
Distance	Gain $_{t-1}$	**(-)**	**(-)**	**(-)**	**(-)**			
	Loss $_{t-1}$		**(+)**				**(+)**	
RideTime	Gain $_{t-1}$			**(+)**				
	Loss $_{t-1}$							**(+)**
Satisfy$_{t-1}$							**(+)**	
Constant		**(+)**	**(+)**	**(+)**	**(+)**	**(+)**	**(+)**	**(+)**



(a) Gender



(b) Distance_Gain



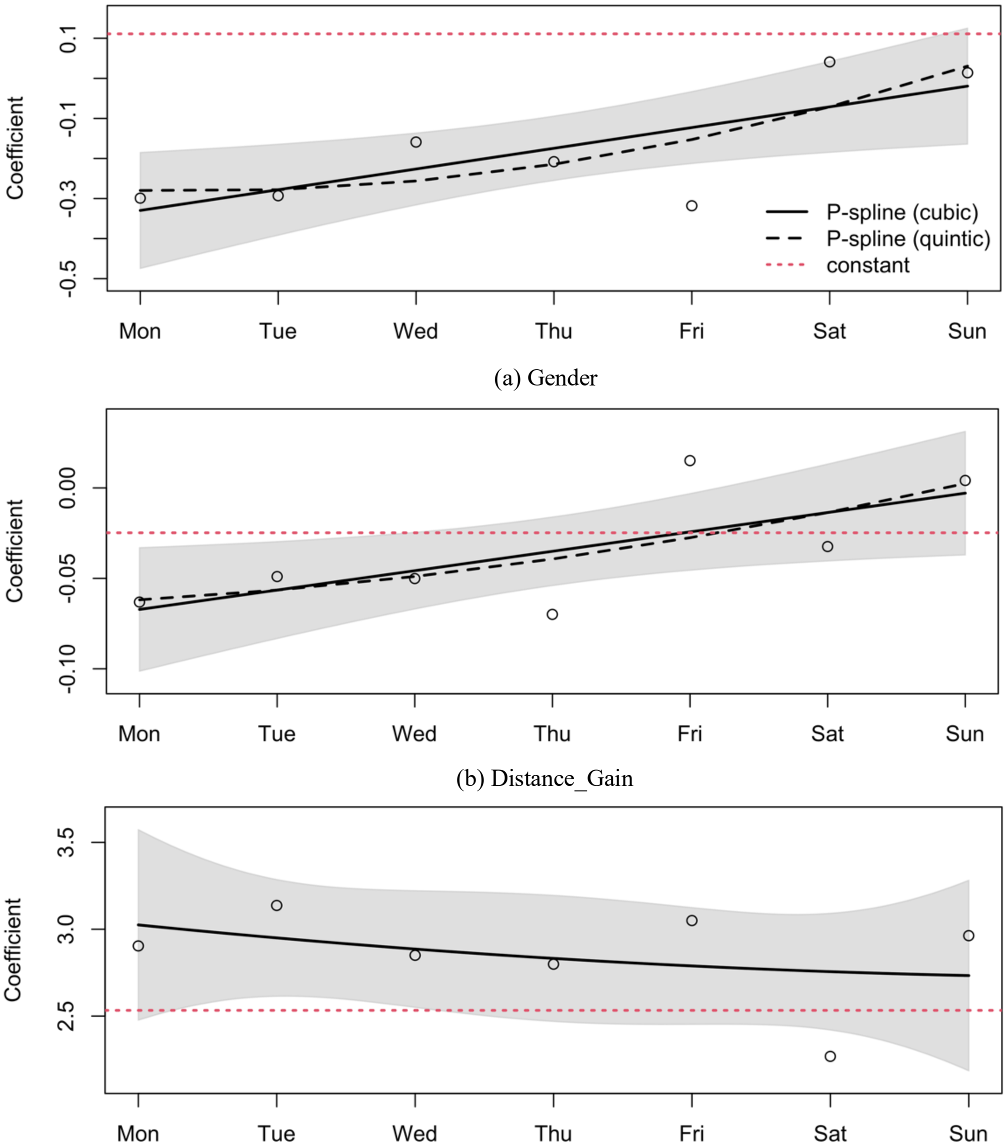
(c) Constant

Figure Appendix 2. P-spline fits for Usage Interval parameters by days of the week

From the figure, it can be seen that being men have a larger usage interval on weekdays and women have a larger usage interval on weekends. This suggests that women use the service more frequently on weekdays and men use it more frequently on weekends. The usage interval influenced by being a man on weekdays decreases as the days approach the weekend. Also, the distance gained with respect to described expectations is negatively significant with the usage interval. That is, the greater the distance gained from expectations, the more frequently the service is used. This tendency is more heavily influenced on Mondays to Thursdays than on Fridays to Sundays.

Then, the coverage of the 95% Bayesian confidence interval (CI) for each smoothing was calculated using Eq. (44). Figure Appendix 3 shows the estimated functional relationship as well as the simulated coverage of the 95% Bayesian CIs. To add, the summary of fit of the estimated splines is summarized in Tables Appendix 3 and 4, each showing the approximate significance of parametric and nonparametric effects, respectively. Statistical inference is conducted via (approximate) frequentist chi-square tests using the Bayesian interpretation of a smoothing spline (Nychka, 1988; Wahba, 1983).

It is noticeable from the results that for both covariates of $Gender$ and $Distance\_Gain$, as well as the constant, the coefficient that is estimated from the generic model do not fall completely within the Bayesian CIs of the cubic spline. This suggests that the nonparametric model should be preferred in fitting the given coefficients that vary with time. Nonetheless, because all nonparametric effects of the covariates are insignificant, this indicates that that the linear effects need to be considered. That is because the time variances are caused by long-term fluctuations. There are no short-term trends in the effect of covariates on the usage intervals of ride-hailing platform's customers.



(a) Gender

(b) Distance_Gain

(c) Constant

Figure Appendix 3. P-spline CIs for Usage Interval parameters by days of the week

Table Appendix 3. Parametric effects from P-spline fits of usage interval parameters by days

Variable	Parameter	Cubic P-Spline				Quintic P-Spline			
		Estimate	Std. Error	t-value	Pr(>|t|)	Estimate	Std. Error	t-value	Pr(>|t|)
Gender	(Intercept)	-0.1746	0.0409	-4.267	0.0080**	-0.1845	0.0441	-4.19	0.0139*
	x	0.3103	0.1227	2.528	0.0527.	0.3103	0.1271	2.44	0.0711.
	x^2	-	-	-	-	0.7183	0.8803	0.82	0.46
$Distance\_Gain_{t-1}$	(Intercept)	-0.0350	0.0097	-3.63	0.0151*	-0.0361	0.0111	-3.26	0.0310*
	x	0.0643	0.0290	2.22	0.0772.	0.0643	0.0319	2.02	0.1141
	x^2	-	-	-	-	0.0769	0.2211	0.35	0.7455
Constant	(Intercept)	-	-	-	-	2.8478	0.1261	22.59	2.275e-05***
	x	-	-	-	-	-0.2916	0.3634	-0.80	4.672e-01
	x^2	-	-	-	-	0.3763	2.5178	0.15	8.884e-01

*** p<0.001, ** p<0.01, * p<0.05, . p<0.1

Table Appendix 4. Nonparametric effects from P-spline fits of usage interval parameters by days

Variable	Parameter	Cubic P-Spline					Quintic P-Spline				
		DF	Sums of Squares	Mean Squares	F-value	Pr(>F)	DF	Sums of Squares	Mean Squares	F-value	Pr(>F)
Gender	$f(x)$	3.496e-05	3.144e-07	0.0090	0.7675	0.4211	3.023e-07	1.385e-09	0.0046	0.36	0.5784
	Residuals	5.000	5.859e-02	0.0117	-	-	4.000	5.023e-02	0.0126	-	-
Distance_Gain$_{t-1}$	$f(x)$	8.535e-07	1.073e-10	0.0001	0.1925	0.6791	7.379e-09	5.574e-13	7.554e-05	0.10	0.7728
	Residuals	5.000	3.253e-03	0.0007	-	-	4.000	3.167e-03	7.919e-04	-	-
Constant	$f(x)$	-	-	-	-	-	7.379e-09	5.476e-10	0.0742	0.7225	0.4432
	Residuals	-	-	-	-	-	4.000	4.109e-01	0.1027	-	-

*** p<0.001, ** p<0.01, * p<0.05, . p<0

Following, the same empirical model used to analyze the stream-of-time effects of covariates on the total service use by the times of the day was used to observe stream-of-time effects of the total service use ($NbrUse$) by the days of the week as our interest is to identify the influence of gains and losses of specified covariates in consumers' total use of service. The estimation results for total usage are presented in Table Appendix 5, and the significant variables for each time of the day are again summarized in Table Appendix 6.

The results reveal that the coefficients for covariates of Age and $WaitTime$ were consistently significant on most days of the week, followed by covariates of $Distance\_Gain$ and $SatisfyUse$ that were only significant for three days in a week. Coefficients for covariates that were consistently significant and varies by day of the week were fit using both the cubic and quintic P-spline functions as in Figure Appendix 4.

From the figure, it can be seen that overall, the older the person, the greater the total service use of ride-hailing platform. However, such influence of age on the total service use decreases as the weekend approaches. The quintic spline additionally shows that such decrease is most sharp from Monday to Thursday, and then the level of decrease diminishes. Also, it was revealed that as the wait time decreases, the total service use increases. Such level of increase diminishes until Wednesday and Thursday and then gradually increases again. This suggests that the effect of wait time reduction on the total service use is most effective on Saturday, Sunday, and Monday, and less effective on other days of the week.

Table Appendix 5. Estimation results for total usage by days

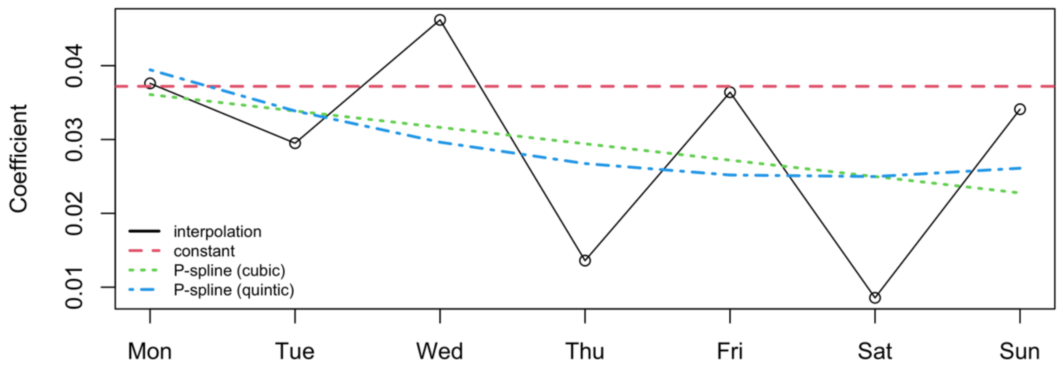
VARIABLES		Monday	Tuesday	Wednesday	Thursday	Friday	Saturday	Sunday
Age		**0.0376*** **	**0.0295*** **	**0.0462*** **	0.0136	**0.0364*** **	0.00856	**0.0341*** **
		(0.00784)	**(0.00886)**	**(0.00793)**	(0.00909)	**(0.00792)**	(0.00722)	**(0.0132)**
Gender		0.149	**-0.265***	-0.140	-0.0479	-0.211	**-0.372*** **	**-0.840*** **
		(0.142)	**(0.144)**	(0.127)	(0.145)	(0.134)	**(0.139)**	**(0.224)**
WaitTime		**-0.000347***	**-0.000215**	**-0.000528*** **	**-0.000472*** **	-0.000146	**0.000226***	**-0.00117*** **
		(0.000210)	**(0.000168)**	**(0.000186)**	**(0.000183)**	(0.000166)	**(0.000136)**	**(0.000306)**
Distance	Gain	0.00643	0.0371	-0.0109	0.0219	**0.105***	**0.0561***	0.179
		(0.0375)	(0.0633)	(0.0470)	(0.0416)	**(0.0583)**	**(0.0334)**	(0.119)
	Loss	0.0598	0.0225	0.0335	0.0450	0.0124	**0.0134***	0.0274
		(0.0378)	(0.0216)	(0.0400)	(0.0301)	(0.0232)	**(0.00687)**	(0.137)
RideTime	Gain	**-0.00129** **	-0.000536	**-0.000660***	-0.000192	**-0.000632***	-0.000164	-0.000690
		(0.000562)	(0.000529)	**(0.000386)**	(0.000368)	**(0.000377)**	(0.000295)	(0.000681)
	Loss	-0.000287	-0.000335	**-0.000865** **	-0.000192	0.000337	-3.19e-05	**-0.00305*** **
		(0.000211)	(0.000410)	**(0.000352)**	(0.000354)	(0.000374)	(0.000389)	**(0.00107)**
Cost	Gain	0.000131	**0.000233***	0.000156	-4.98e-05	-7.26e-05	**-0.000128** **	-0.000164
		(0.000135)	**(0.000129)**	(0.000123)	(8.72e-05)	(9.49e-05)	**(5.63e-05)**	(0.000131)
	Loss	-1.34e-05	-5.27e-05	-1.57e-05	6.71e-07	**-0.000193** **	-5.52e-05	-2.97e-05
		(1.36e-05)	(7.32e-05)	(1.47e-05)	(6.30e-05)	**(8.29e-05)**	(3.95e-05)	(0.000154)
Satisfy		0.457	0.465	**-1.033*** **	0.831	**0.877***	0.519	**-0.979** **

	(0.398)	(0.440)	**(0.379)**	(1.075)	**(0.491)**	(0.453)	**(0.417)**
Constant	-0.236	-0.133	**1.099****	0.243	-0.417	0.287	**2.204*****
	(0.534)	(0.596)	**(0.459)**	(1.115)	(0.656)	(0.522)	**(0.701)**
Observations	206	188	237	199	269	189	100
/lnalpha	**-0.374*****	**-0.579*****	**-0.471*****	**-0.451*****	**-0.176***	**-0.882*****	**-0.370****
	(0.122)	**(0.144)**	**(0.120)**	**(0.132)**	**(0.1000)**	**(0.173)**	**(0.175)**
Log-Likelihood	-532.4360	-430.0043	-603.9028	-485.3961	-701.7730	-381.3279	-253.2531

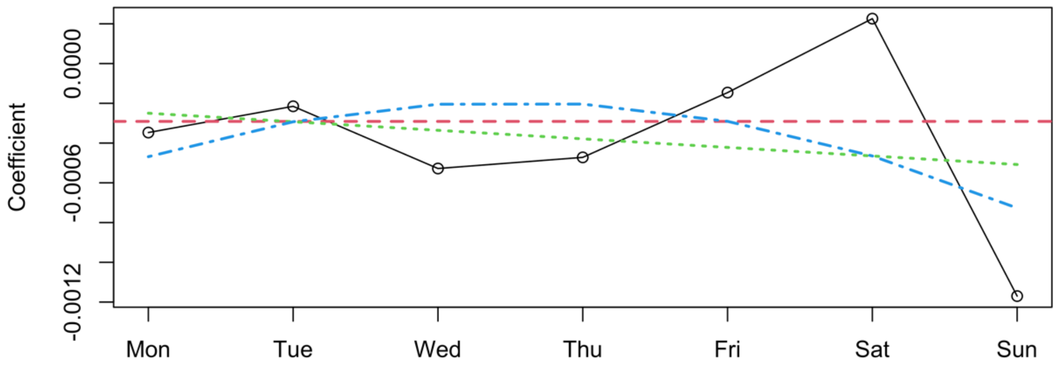
*** p<0.01, ** p<0.05, * p<0.1; Standard errors are in parenthesis

Table Appendix 6. Summary of significant variables for total usage by days

VARIABLES		Monday	Tuesday	Wednesday	Thursday	Friday	Saturday	Sunday
Age		(+)	(+)	(+)		(+)		(+)
Gender			(-)				(-)	(-)
WaitTime		(-)	(-)	(-)	(-)		(+)	(-)
Distance	Gain					(+)	(+)	
	Loss						(+)	
RideTime	Gain	(-)		(-)		(-)		
	Loss			(-)				(-)
Cost	Gain		(+)				(-)	
	Loss					(-)		
Satisfy				(-)		(+)		(-)
Constant				(+)				(+)



(a) Age

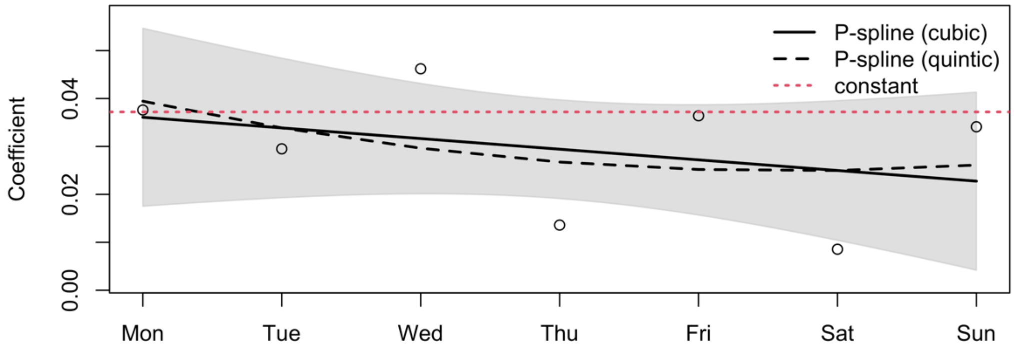


(b) WaitTime

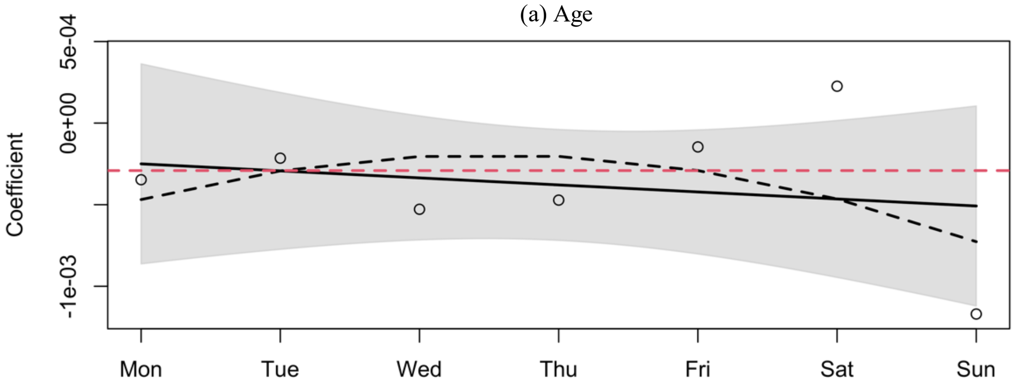
Figure Appendix 4. P-spline fits for Total Usage parameters by days of the week

Then, the coverage of the 95% Bayesian confidence interval (CI) for each smoothing was calculated using Eq. (44). Figure Appendix 5 shows the estimated functional relationship as well as the simulated coverage of the 95% Bayesian CIs. To add, the summary of fit of the estimated splines is summarized in Tables Appendix 7 and 8, each showing the approximate significance of parametric and nonparametric effects, respectively.

Putting the results from Figure Appendix 5 together with Tables Appendix 7 and 8, it is noticeable that for both covariates of Age and $WaitTime$, the coefficients that are estimated from the generic model completely falls within the Bayesian CIs of the cubic spline. This indicates that all models sufficiently explain the data, with the nonparametric model showing the long-term variance of time. Likewise, because all nonparametric effects of the covariates are insignificant, this indicates that that the linear effects need to be considered. There are no short-term trends in the effect of covariates on the total usage of ride-hailing platform's customers, and their preferences are rather consistent with time.



(a) Age



(b) WaitTime

Figure Appendix 5. P-spline CIs for Total Usage parameters by days of the week

Table Appendix 7. Parametric effects from P-spline fits of total usage parameters by days

		Cubic P-Spline				Quintic P-Spline			
Variable	Parameter	Estimate	Std. Error	t-value	Pr(>$\|t\|$)	Estimate	Std. Error	t-value	Pr(>$\|t\|$)
Age	(Intercept)	0.0294	0.0053	5.60	0.0025**	0.0288	0.0060	4.80	0.0087**
	x	-0.0133	0.0158	-0.85	0.4366	-0.0133	0.0173	-0.77	0.4836
	x^2	-	-	-	-	0.0483	0.1197	0.40	0.7075
WaitTime	(Intercept)	-0.0004	0.0002	-2.18	0.0807.	-0.0003	0.0002	-1.80	0.1456
	x	-0.0003	0.0005	-0.50	0.6408	-0.0003	0.0005	-0.48	0.6549
	x^2	-	-	-	-	-0.0032	0.0037	-0.85	0.4437

*** p<0.001, ** p<0.01, * p<0.05, . p<0.1

Table Appendix 8. Nonparametric effects from P-spline fits of total usage parameters by days

Variable	Parameter	Cubic P-Spline DF	Sums of Squares	Mean Squares	F-value	Pr(>F)	Quintic P-Spline DF	Sums of Squares	Mean Squares	F-value	Pr(>F)
Age	$f(x)$	8.535e-07	3.653e-11	4.281e-05	0.22	0.6577	7.379e-09	8.505e-13	1.115e-04	0.50	0.5199
	Residuals	5.000	9.664e-04	1.933e-04	-	-	4.000	9.287e-04	2.322e-04	-	-
WaitTime	$f(x)$	3.496e-07	5.769e-14	1.650e-07	0.78	0.4166	3.023e-09	1.325e-15	4.384e-07	1.97	0.2335
	Residuals	5.000	1.053e-06	2.106e-07	-	-	4.000	8.921e-07	2.230e-07	-	-

*** p<0.001, ** p<0.01, * p<0.05, . p<0.1

Lastly, the same empirical model used to analyze the stream-of-time effects of covariates on service satisfaction of consumers in the Macaron ride-hailing platform service by the times of the day was used to observe stream-of-time effects of the service satisfaction ($Satisfy$) that varies by days of the week. Our interest is to identify the influence of gains and losses of specified covariates in consumers' service satisfaction. The estimation results for service satisfaction are presented in Table Appendix 9, and the significant variables for each time of the day are again summarized in Table Appendix 10.

The results reveal that the coefficients for covariates of $Gender$ and $RideTime\_Loss$ were consistently significant on each day of the week, while the covariate of $Satisfy_{t-1}$, indicating whether the consumer experience of prior ride-hailing use was satisfactory, was significant on all days. Covariates with regard to cost were almost always insignificant. The estimated paths for coefficients of covariates that were consistently significant and varies by the day of the week were depicted using both the cubic and quintic P-spline functions as shown in Figure Appendix 6.

From the figure, we obtain rather smooth curves for $Gender$ and $Satisfy_{t-1}$, suggesting long-term trends in their time-varying influence on satisfaction. Specifically, it can be seen that overall, men are more satisfied with the Macaron Taxi ride-hailing service than women. The level of satisfaction influenced by being a man decreases from Monday to Thursday, and again increases from Friday to Sunday, suggesting that on Thursdays and Fridays, consumers' level of satisfaction is least influenced by gender. It should be noted that although the constant estimate from the generic model seems to imply that women are more satisfied, the coefficient of gender was insignificant in that model – the constant graph is drawn just for reference, not for interpretation.

Table Appendix 9. Estimation results for satisfaction by days

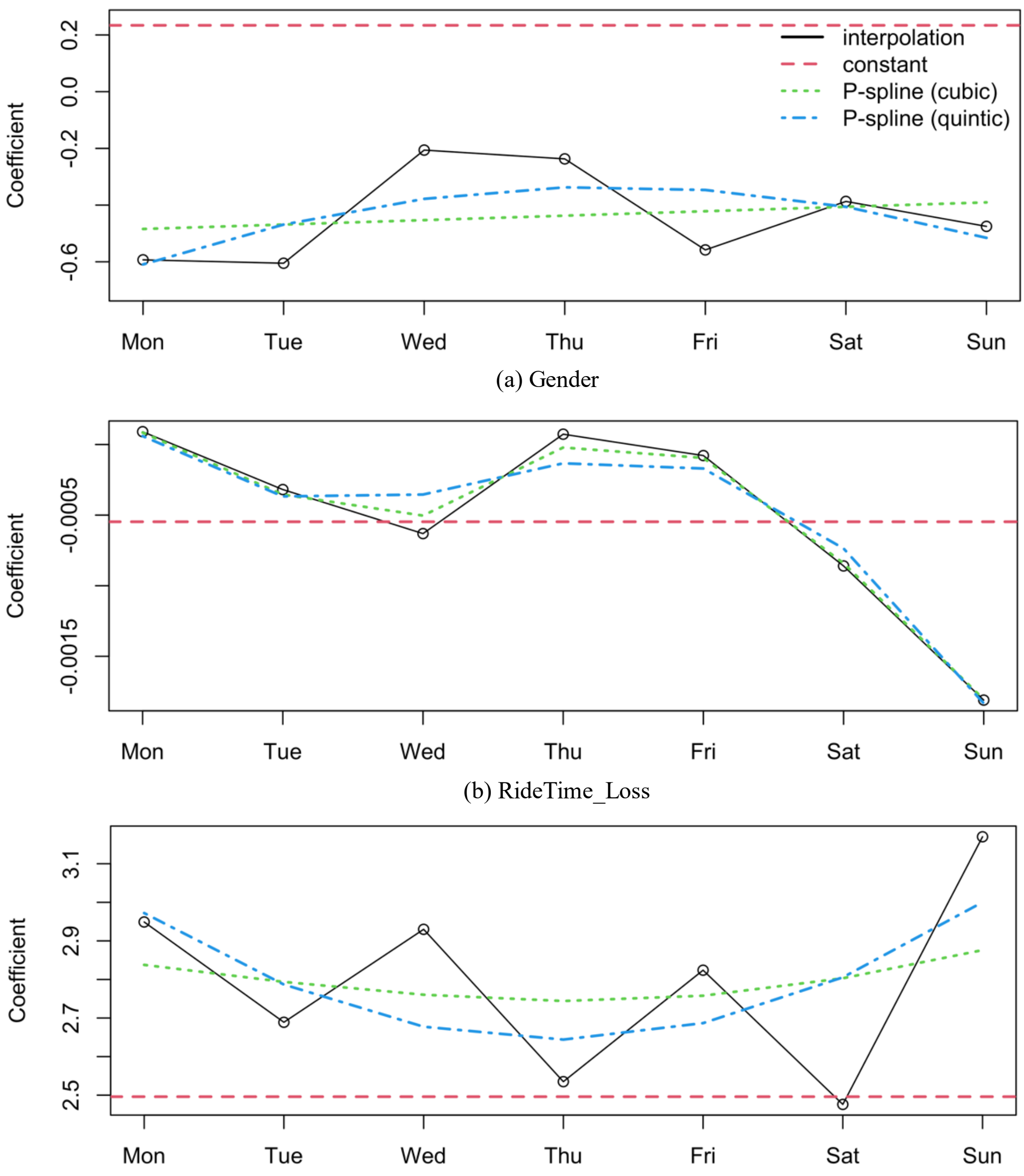
VARIABLES		Monday	Tuesday	Wednesday	Thursday	Friday	Saturday	Sunday
Age		**-0.0177****	**-0.0169***	0.00213	-0.0120	-0.00973	-0.0144	-0.00571
		(0.00880)	**(0.00898)**	(0.00830)	(0.00930)	(0.00750)	(0.00907)	(0.0131)
Gender		**-0.593*****	**-0.605*****	-0.206	-0.237	**-0.558*****	**-0.387****	**-0.475***
		(0.169)	**(0.174)**	(0.157)	(0.167)	**(0.145)**	**(0.172)**	**(0.251)**
WaitTime		0.000340	**0.000492****	0.000153	**0.000335***	0.000194	**0.000407****	0.000364
		(0.000229)	**(0.000211)**	(0.000184)	**(0.000191)**	(0.000181)	**(0.000190)**	(0.000309)
Distance	Gain	-0.00777	-0.0713	-0.0416	**0.0730***	-0.0337	0.0350	-0.0533
		(0.0525)	(0.0639)	(0.0373)	**(0.0423)**	(0.0337)	(0.0418)	(0.0822)
	Loss	-0.0141	**0.0556***	-0.00802	0.00630	-0.00804	**0.0349****	0.00293
		(0.0127)	**(0.0289)**	(0.0278)	(0.0405)	(0.0251)	**(0.0149)**	(0.0854)
RideTime	Gain	0.000155	-0.000304	0.000563	-0.000217	-1.28e-05	0.000171	**-0.00151****
		(0.000483)	(0.000464)	(0.000344)	(0.000373)	(0.000312)	(0.000375)	**(0.000748)**
	Loss	**9.07e-05**	**-0.000319**	**-0.000631****	**7.30e-05**	**-7.84e-05**	**-0.000860*****	**-0.00181****
		(0.000235)	**(0.000304)**	**(0.000297)**	**(0.000309)**	**(0.000220)**	**(0.000331)**	**(0.000902)**
Cost	Gain	0.000139	**0.000233****	-2.49e-05	-0.000151	-2.24e-05	-2.24e-05	0.000137
		(0.000120)	**(0.000117)**	(7.55e-05)	(0.000104)	(9.25e-05)	(6.19e-05)	(0.000147)
	Loss	2.17e-05	-2.50e-05	3.93e-06	-9.67e-06	-1.37e-05	1.89e-05	-1.78e-05
		(2.63e-05)	(6.57e-05)	(1.85e-05)	(4.47e-05)	(3.30e-05)	(4.90e-05)	(9.48e-05)
Satisfy $_{t-1}$		**2.949*****	**2.689*****	**2.930*****	**2.535*****	**2.824*****	**2.476*****	**3.170*****

	(0.357)	**(0.375)**	**(0.338)**	**(0.350)**	**(0.308)**	**(0.374)**	**(0.502)**
Observations	1,190	1,405	1,471	1,354	1,503	1,130	511
/cut1	**-4.543*** **	**-5.076*** **	**-3.935*** **	**-6.413*** **	**-4.804*** **	**-4.906*** **	**-3.492*** **
	(0.587)	**(0.638)**	**(0.571)**	**(1.101)**	**(0.564)**	**(0.625)**	**(0.765)**
/cut2	**2.514*** **	**2.603*** **	**3.405*** **	**2.789*** **	**2.567*** **	**2.490*** **	**2.795*** **
	(0.453)	**(0.489)**	**(0.441)**	**(0.470)**	**(0.388)**	**(0.430)**	**(0.672)**
Log-Likelihood	-533.9838	-537.9248	-635.8605	-539.9025	-684.3139	-498.1884	-249.3963

*** p<0.01, ** p<0.05, * p<0.1; Standard errors are in parenthesis

Table Appendix 10. Summary of significant variables for satisfaction by days

VARIABLES		Monday	Tuesday	Wednesday	Thursday	Friday	Saturday	Sunday
Age		(-)	(-)					
Gender		(-)	(-)			(-)	(-)	(-)
WaitTime			(+)		(+)		(+)	
Distance	Gain				(+)			
	Loss		(+)				(+)	
RideTime	Gain							(-)
	Loss	(+)	(-)	(-)	(+)	(-)	(-)	(-)
Cost	Gain		(+)					
	Loss							
Satisfy $_{t-1}$		(+)	(+)	(+)	(+)	(+)	(+)	(+)



(a) Gender

(b) RideTime_Loss

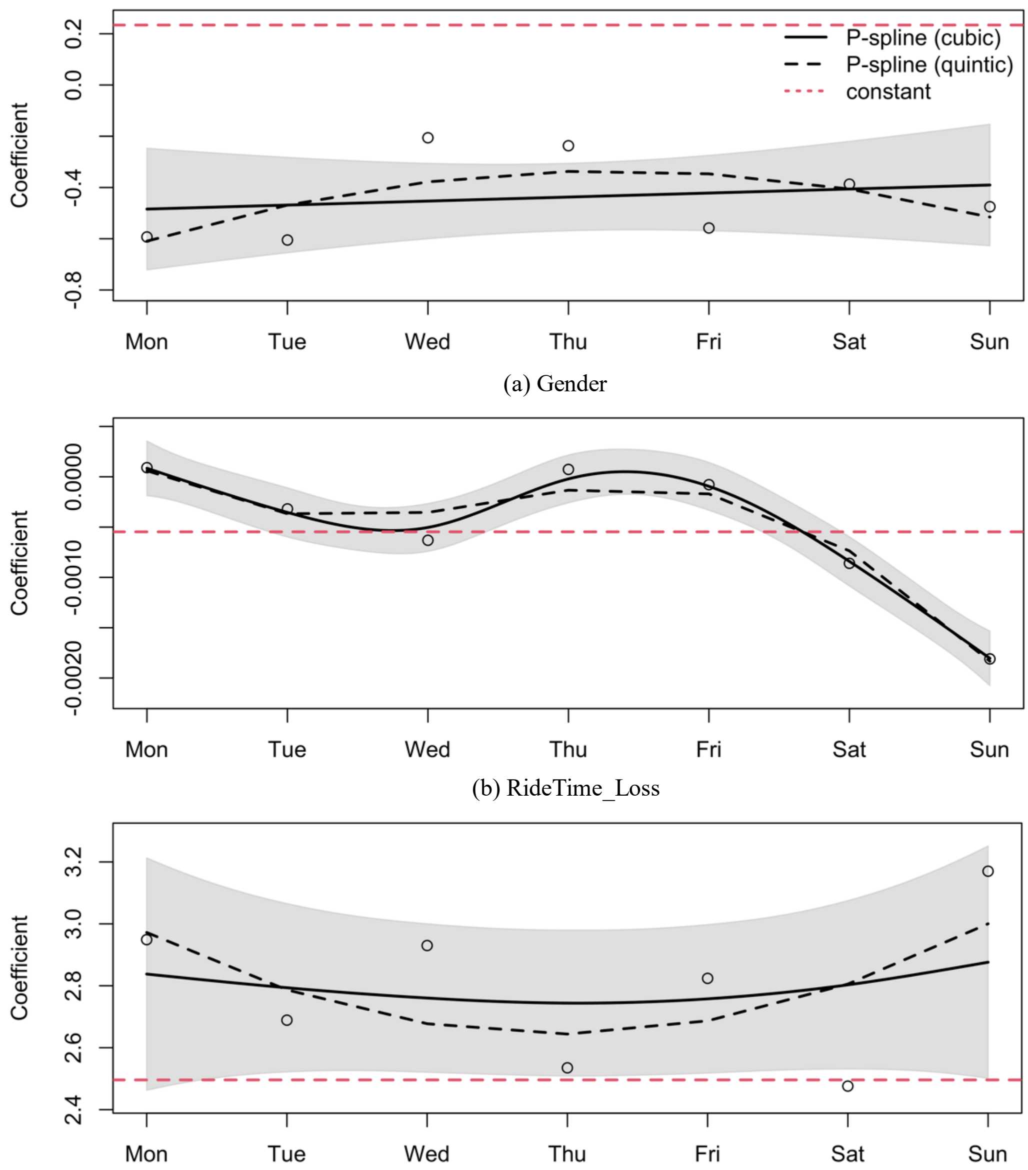
(c) Satisfy_{t-1}

Figure Appendix 6. P-spline fits for Satisfaction parameters by days of the week

The results also reveal that as the consumer's satisfaction in the previous use of service positively influences the service satisfaction of the current use. Such level of influence diminishes until Thursday and then increase again. This suggest that the effect of satisfaction in the previous use of service is of least influence on the current satisfaction of the service on Thursday.

Comparatively, the time-varying coefficients of $RideTime\_Loss$ suggest short-term trend. While the decrease in loss of ride time relative to described expectations yield increase in service satisfaction, such level of increase in satisfaction amplifies from Monday to Wednesday and lessens back to the level of that on Monday by Thursday. Then, as the day approaches weekend, consumers are again more sensitively influenced by the loss in ride time, where the increase of marginal gain in service satisfaction becomes sharper from Friday to Sunday than from Thursday to Friday. The tendency also suggests that the reduction of gap between the expected ride time as suggested by the platform system and the actual ride time that the consumer experiences will most increase consumers' service satisfaction on weekends compared to other days of the week.

Then, the coverage of the 95% Bayesian confidence interval (CI) for each smoothing was calculated. Figure Appendix 7 shows the estimated functional relationship as well as the simulated coverage of the 95% Bayesian CIs. To add, the summary of fit of the estimated splines is summarized in Tables Appendix 11 and 12, each showing the approximate significance of parametric and nonparametric effects, respectively.



(a) Gender

(b) RideTime_Loss

(c) Satisfy_{t-1}

Figure Appendix 7. P-spline CIs for Satisfaction parameters by days of the week

Putting the results from Figure Appendix 7 together with Tables Appendix 11 and 12, it is noticeable that for both all covariates, constant coefficients that is estimated from the generic model do not fall completely within the Bayesian CIs of the cubic spline estimations, suggesting that the nonparametric model should be preferred in fitting the given coefficients that vary with time. Nonetheless, for covariates of $Gender$ and $Satisfy_{t-1}$, all nonparametric effects are insignificant, indicating that the linear effects involved in the smooth need to be considered, again representing long-term trend of time-varying coefficients. Contrastingly, because the nonparametric part of the $RideTime\_Loss$ is significant, which refers to the nonlinearity beyond the linear/parametric part of the smooth, a linear effect of the covariate is not supported by the data.

Table Appendix 11. Parametric effects from P-spline fits of satisfaction parameters by days

		Cubic P-Spline				Quintic P-Spline			
Variable	Parameter	Estimate	Std. Error	t-value	Pr(>$\|t\|$)	Estimate	Std. Error	t-value	Pr(>$\|t\|$)
Gender	(Intercept)	-0.4373	0.0671	-6.52	0.0013**	-0.4123	0.0638	-6.46	0.0030**
	x	0.0939	0.2013	0.4662	0.6607	0.0939	0.1839	0.51	0.6366
	x^2	-	-	-	-	-1.8000	1.2737	-1.41	0.2305
RideTime _Loss	(Intercept)	-0.0004	5.777e-05	-7.59	0.0453*	-0.0004	9.317e-05	-4.635	0.0245*
	x	-0.0019	1.957e-04	-9.65	0.0329*	-0.0019	3.202e-04	-5.912	0.0134*
	x^2	-	-	-	-	-0.0041	3.655e-03	-1.129	0.3503
$Satisfy_{t-1}$	(Intercept)	2.7843	0.0970	28.70	2.451e-06***	2.2781	0.0968	28.50	9.016e-06***
	x	0.0383	0.2916	0.13	9.011e-01	0.0281	0.2789	0.10	9.247e-01
	x^2	-	-	-	-	2.7369	1.9323	1.42	2.296e-01

*** p<0.001, ** p<0.01, * p<0.05, . p<0.1

Table Appendix 12. Nonparametric effects from P-spline fits of satisfaction parameters by days

Variable	Parameter	Cubic P-Spline DF	Cubic P-Spline Sums of Squares	Cubic P-Spline Mean Squares	Cubic P-Spline F-value	Cubic P-Spline Pr(>F)	Quintic P-Spline DF	Quintic P-Spline Sums of Squares	Quintic P-Spline Mean Squares	Quintic P-Spline F-value	Quintic P-Spline Pr(>F)
Gender	$f(x)$	3.496e-05	1.788e-06	0.0511	1.62	0.2588	3.023e-07	4.272e-09	0.0141	0.54	0.5041
	Residuals	5.000e+00	1.577e-01	0.0315	-	-	4.000e+00	1.052e-01	0.0263	-	-
RideTime _Loss	$f(x)$	3.649	2.261e-06	3.456e-07	17.24	0.1146	1.335	5.850e-07	4.381e-07	7.957	0.0761.
	Residuals	1.351	2.708e-08	2.005e-08	-	-	2.665	1.467e-07	5.505e-08	-	-
Satisfy_{t-1}	$f(x)$	0.430	0.0414	0.0964	1.50	0.2800	7.397e-09	3.489e-10	0.0473	0.78	0.4266
	Residuals	4.570	0.2935	0.0642	-	-	4.000	2.420e-01	0.0605	-	-

*** p<0.001, ** p<0.01, * p<0.05, . p<0.1

Abstract (Korean)

서비스에 대한 고객 충성도를 야기하고 유지하기 위해서는 지속적으로 소비자를 만족시켜야 하고, 소비자의 서비스 재이용 여부는 해당 서비스와 관련하여 축적된 소비자의 이용 경험에 따라 달라진다는 것은 마케팅 문헌에서 익히 알려져 있다. 그러나 기존 연구는 재이용이 빈번한 플랫폼 서비스에서의 사용자 행태를 분석하는데 있어 정적 효용 모형을 사용하므로, 서비스의 지속 사용에 따른 시간 변동 효과를 적절하게 보이지 못하고 있다. 또한 소비자의 지속적인 사용에 따른 고객의 서비스에 대한 기대치 및 인식이 변화할 수 있음을 반영하지 못한다. 본 연구에서는 동적 효용 모형을 채택함으로써 플랫폼 사용자가 서비스 이용 경험에 기반하여 서비스에 대한 기대치 및 인식을 조정할 수 있음을 반영하고, 서비스에 대한 기대치와 실제 경험의 차이가 서비스의 만족도에 어떻게 영향을 미치는지를 알아보고자 한다. 또한 반모수 모델링을 통해 소비자의 서비스 이용 행태에서의 공변량의 시간적 특성을 알아본다. 분석 결과, 서비스에 대한 '서비스 격차' 및 '인식 격차'는 서비스 만족도에 영향을 미치며, 그 영향 수준은 경험이 누적됨에 따라 변화함을 알 수 있었다. 또한, 누적된 경험에 기반하여 조정된 소비자의 서비스에 대한 인식과 실제 서비스 이용 경험 간의 차이가 서비스 만족에 대한 경험 누적 효과를 가장 잘 설명함을 알 수 있었다. 마지막으로 소비자의 서비스 이용에 있어 계절적 특성이 있음을 알 수 있었다. 이에 마케팅 관점에서 공변량에 대한 시간적 효과

를 반영하지 못하면 소비자의 행동 변화를 잘못 감지할 가능성이 있으므로, 마케팅 전략 수립에 있어 서비스 재이용에 따른 특성 및 계절성을 적절히 반영할 필요가 있음을 알 수 있었다.

주요어 : 소비자 이용행태, 반복사용, 반모수 모델링, 기대치 조정, 시간적 효과, 플랫폼 서비스
학 번 : 2017-32883