Loyalty Currency and Mental Accounting: Do Consumers Treat Points Like Money?

Freddy Lim
Decision Sciences, INSEAD, 1 Ayer Rajah Avenue, Singapore 138676. freddy.lim@insead.edu

So Yeon Chun
Technology and Operations Management, INSEAD, 1 Ayer Rajah Avenue, Singapore 138676. soyeon.chun@insead.edu

Ville Satopää
Technology and Operations Management, INSEAD, Boulevard de Constance, Fontainebleau, France 77305.
ville.satopaa@insead.edu

Problem definition: Loyalty programs have greatly expanded in scale and scope, and loyalty points issued by firms serve as a new form of currency alongside the traditional currency of money. In this paper, we study how consumers decide to pay with points or money for a purchase and how these decisions are affected by consumers' points earning characteristics.

Methodology/results: We develop a model of consumers' payment choices and estimate it on proprietary loyalty program data from a major U.S. airline company using a hierarchical Bayesian framework. Our results demonstrate that mental accounting, the subjective perceived value of points, and the reference exchange rate play important roles in consumers' payment choices. Moreover, the primary points earning source and the total earning level are jointly associated with consumers' attitudes toward points and money: Consumers who earn many points and mostly with the focal firm tend to value points more than money, while those who earn few points or mostly through a co-branded credit card tend to value money more than points. To better understand heterogeneity in consumers' attitudes toward points, we propose a probabilistic segmentation of consumers and identify four behavioral segments with distinctive characteristics. Through counterfactual analysis, we demonstrate how a firm can implement money and point pricing policies to optimally target and influence consumers' payment choices.

Managerial implications: It is important for firms to understand how consumers think about points and decide to pay with points or money as it affects firms' cash flows, profitability, and consumer loyalty and engagement. We demonstrate that firms can design optimal price targeting policies by utilizing the significant level of heterogeneity in consumers' attitudes toward points. Furthermore, our results indicate that firms that are expanding their partnership networks and offering more points earning sources should consider the impact that this would have on consumers' mental accounting of money and points, and consequently their payment choices.

Key words: payment choice; mental accounting; the reference effect; point currency; loyalty program

1. Introduction

Many loyalty programs in several industries award their consumers points for their purchases. Given that these points can be redeemed later for additional products or services, they effectively serve as a new currency (Chun et al. 2020).

Recent years have witnessed a rapid expansion in the number, size, and scope of consumer loyalty programs. In fact, the average consumer in the United States belongs to more than ten different loyalty programs (Hsu 2019) and a large volume of loyalty points are earned and spent each year. For instance, consumers of American Airlines' loyalty program earned 315 billion points/miles in 2015 (American Airlines 2015, 2019) and passengers of Southwest Airlines redeemed 10.7 million award flights in 2019, which represents 14.1% of revenue passenger miles (paid passenger flown miles) and a 330% increase from the 3.2 million award flight redemptions made in 2010 (Southwest 2010, 2019). According to *The Economist*, already in 2005 an estimated 14 trillion frequent flyer miles worth more than $700 billion were in circulation worldwide (Economist 2005). In addition, for the first time in history, many airlines have used their loyalty programs and point currencies as collateral to raise much needed loans during the COVID-19 pandemic when global air travel was decimated (Chun and de Boer 2021).

Consumers are often able to choose whether to pay with a traditional national currency such as the U.S. dollar (*money*) or to redeem their loyalty points (*points*) for a purchase.[1] There are several reasons why their choices matter to firms. First, these choices affect firms' short-term cash flow and revenue. When consumers choose to pay with points instead of money, a firm's cash inflow is reduced due to the decreased money sale of products and services. Second, the use of point currency affects consumer loyalty and engagement. Loyalty programs that encourage consumers to participate and redeem their points can improve customer retention and strengthen customer relationships (Carluccio et al. 2021), increase consumers' future purchases with the firm due to the positive emotions evoked when they redeem points for rewards (Bijmolt et al. 2010), and improve the long-term sustainability of the program (Schlangenstein and Bachman 2019). Third, consumers' payment choices also directly impact firms' earnings and profitability (KPMG 2022, Chun et al. 2020). Given that loyalty points represent a promise for future service, their monetary value counts as a liability on the firm's balance sheet. According to a new accounting standard[2] (effective in the United States as of 2018), firms are required to calculate this liability using a "deferred revenue" method: Upon making an initial cash sale for which points are awarded, firms must defer the portion of the sales revenue that corresponds to the value of the awarded points. Although this reduces the reported revenue (and profit) from the sale, firms can recognize the deferred revenues and hence

[1] Recently, some firms have started to offer an option to pay with a mixture of money and points. However, paying fully with money or with points is still the dominant decision choice for consumers. For example, in the airline industry, many airlines including American, Southwest, and Frontier still do not offer a mixed payment option as of December 2023. Other airlines such as United, Delta, and JetBlue have introduced a mixed payment option, but this is not available for certain fares and markets.

[2] IFRS 15 Revenue from Contracts with Customers. For more details, see https://www.ifrs.org/issued-standards/list-of-standards/ifrs-15-revenue-from-contracts-with-customers

increase profits later on when the points are redeemed. The sheer magnitude of loyalty programs has turned these liabilities into significant items on the firms' balance sheets. Delta Air Lines, for example, recorded about $7.2 billion of loyalty points liabilities (about 10% of their total liabilities) and about $935 million revenue from redeemed points (about 7% of total passenger revenue) in 2020 (Delta Air Lines 2020). Consequently, a small change in the way that consumers use their points can have a first-order impact on the firms' profitabilities. Despite this high economic and financial impact of loyalty point redemptions, consumers' payment choice behaviors are not well understood.

In addition to growing in scale over the years, loyalty programs have increased the ways in which points can be earned. In particular, there has been a dramatic growth in the number of co-branded credit card[3] offerings and the amount of loyalty points earned through the use of co-branded credit cards. Among several industries, co-branded card partnerships have become both an important source of revenue for the firms and an important source of points earning for consumers. For example, in 2019 American Airlines realized ancillary revenue of $2.4 billion from selling miles (mainly to banks), a considerable amount compared to their total operating income of $3.1 billion in the same year (American Airlines 2015, 2019). The expansion of the number of points earned through co-branded credit card partnerships and their financial significance to the firms highlight the importance of understanding how consumers' attitudes toward point currency and their payment choices may be affected by their points earning sources.

Even though there is abundant literature on loyalty programs, past work has largely focused on how effective the loyalty programs are in generating sales revenue. This leaves several important research questions unanswered: How do consumers decide whether to pay with points or money? Do consumers treat points like money? Specifically, how do they mentally account for points compared to money and how does such mental accounting depend on whether the points were earned through spending with the focal firm or through spending with co-branded credit cards? Finally, based on the understanding of consumers' attitudes toward points and their payment choices, how can firms set the money and point prices to target different consumers with different pricing policies?

To address these questions, we develop a model of consumer payment choices. The model incorporates several behavioral components such as mental accounting (i.e., weighing the utility of paying with points versus money based on separate utility functions), reference effects (i.e., comparing the offered points exchange rate to a reference points exchange rate), and points earning source effects (i.e., earning points through the focal firm versus co-branded credit cards). We estimate the model on proprietary loyalty program transaction data from a major U.S. airline using a hierarchical Bayesian framework, which allows us to capture the heterogeneity among consumers.

[3] A co-branded credit card partnership between the focal firm and a financial institution issuing credit cards allows consumers who spend on the credit card to earn the focal firm's loyalty points.

1.1. Summary of findings

We demonstrate that mental accounting, the reference exchange rate, the subjective perceived value of points, and the points earning sources play important roles and lead to significant improvements in explaining consumers' payment choice behaviors. In particular, consumers tend to use separate (mental account) functions to evaluate the disutility of paying with points and the disutility of paying with money. Consumers assess the offered points exchange rate relative to a reference rate. Furthermore, the reference rate has an asymmetrical effect on consumers' choices: Consumers react considerably to favorable exchange rates ("gains"), but they are relatively insensitive to unfavorable ones ("losses"). Consumers' subjective perceived value of points varies across decision occasions leading to behavioral changes over time. For example, when consumers temporarily earn many bonus points and have higher point balances, their perceived value tends to be lower, which increases the propensity to pay with points instead of money.

Moreover, we show that the total level and primary source of points earning are jointly associated with consumers' attitudes toward points and money: Consumers who earn many points and mostly with the focal firm are more sensitive to losing points and hence value points more than money. In contrast, consumers who earn fewer points or mostly from the use of co-branded credit cards tend to value money more than points. In between these extremes, consumers who either earn many points from diversified sources or earn an average number of points almost entirely with the focal firm tend to treat points more like money.

To better understand the heterogeneity in consumers' attitudes toward points and their payment choice behaviors, we propose a probabilistic k-medoids segmentation approach. This identifies four behavioral segments with distinctive characteristics: Consumers who are more sensitive to money utility (*"money advocates"*), consumers who make no mental distinction between money and points (*"currency impartialists"*), consumers who actively seek out good exchange rate deals (*"point gamers"*), and consumers who highly value points (*"point lovers"*). Our segmentation results demonstrate that even if the different segments of consumers exhibit similar points redemption propensities, the reasons behind their redemptions may differ substantially.

Finally, we perform counterfactual policy simulations to study how the firm can leverage the significant level of heterogeneity in consumers' attitudes toward points and their differences in behavioral characteristics to deploy optimal point and money pricing policies to influence consumers' payment choices. There are two main insights. First, when the firm implements a point discount policy to incentivize point payment, it increases its revenue from the baseline of no discounts by a larger factor than when it implements a money discount policy to incentivize money payment. This is because consumers in the program are more sensitive to point discounts than to money discounts. Second, even though different consumer segments have similar optimal levels of money discounts that maximize revenues, their optimal levels of point discounts are different, highlighting consumer heterogeneity in attitudes toward points.

2. Literature Review

Our study is mainly related to mental accounting, transaction utility, and reference dependence literature. Thaler (1999) posits that consumers assign spending to specific mental accounts when making individual consumption decisions. For example, consumers may "record" spending transactions using different mental accounts in order to keep track of different types of expenditures (e.g., Soman 2001). We allow consumers to book money and point transactions in separate mental accounts to capture different attitudes that consumers might exhibit toward points and money. Thaler (1985) also presents the reference dependence effects to construct a transaction utility theory where consumers' transaction utilities depend on how much they pay relative to a reference level. Consumers have been consistently found to compare observed prices with benchmark or reference levels before making purchase decisions (for a review, see Mazumdar et al. 2005). However, the perceived transaction utility from a good deal (i.e., when the price is better than the reference level) and a comparably bad deal (i.e., when the price is worse than the reference level) may not be the same (e.g., Erdem et al. 2001). We contribute to this literature by studying such asymmetric effects on consumers' choices to pay with points versus money.

Given that our study demonstrates how a firm can implement efficient operational policies based on consumer behavior, it adds to the branch of operations management literature that incorporates psychological processes to model consumers' decisions. Examples of the behavioral aspects captured in this line of literature include the effect of customer emotions or peer performances and experiences on service agents' behaviors (e.g., Altman et al. 2021, Tan and Netessine 2019, Huckman et al. 2009), reference levels (of price, capacity and quality) and customer regret on sales (e.g., Tereyağoğlu et al. 2018, Martínez-de-Albéniz et al. 2020, Baron et al. 2020, Özer et al. 2020), and customers' perceptions of waiting on the system performance (e.g., Lu et al. 2013, Yu et al. 2017, Buell 2021). By explicitly modeling and estimating consumer heterogeneity, our research also contributes to the growing literature that empirically examines heterogeneous effects (e.g., Keppler et al. 2021, Arslan et al. 2021, Ertekin et al. 2020, Lu et al. 2013).

While the marketing and economics literatures have discussed loyalty programs extensively, past research has focused mainly on the points earning behavior and the effectiveness of loyalty programs in generating sales revenue (e.g., Bijmolt et al. 2010, Breugelmans et al. 2015, Orhun et al. 2021). In contrast, only a few previous studies analyze the factors that influence the consumers' decisions to spend points: Drèze and Nunes (2004) demonstrate that consumers may prefer combined currency payments (money plus points) over single currency payments (either money or points) if they have separate (mental) accounts for money and points and their cost functions include a convex region for one of the currencies. Building upon Drèze and Nunes (2004), Stourm et al. (2015) model mental accounting and redemption choices in a retail loyalty program with fixed points exchange rates, and describe economic, cognitive, and psychological motivations behind the consumers' tendencies to stockpile points. Similarly, Chung (2020) finds that consumers tend to redeem credit card rebates when they can offset a larger proportion of the asking price because the psychological utility gained makes up for the monetary loss from delayed redemptions. Chun and Hamilton (2021) conduct a series of experiments to investigate the behavioral bias that consumers exhibit toward point currency. They show that higher numerosity and variability in points exchange rates reduce consumers' redemption propensities due to the consumers' positive biases on the values of points. We contribute to this literature by holistically capturing mental accounting, the subjective perceived value of points, the reference points exchange rate, and points earning characteristics within a model of consumer payment choice between points and money. Furthermore, we demonstrate how a firm can influence consumers' payment choices with different point and money pricing policies. Additionally, we investigate how points earning sources are associated with consumers' attitudes toward points and money, and hence their payment choice decisions.

Finally, our study relates to the behavioral economics literature that examines how income affects spending in the context of money. In particular, consumers may violate the fundamental assumption of money fungibility by labeling money based on the income type and source. For instance, Epley et al. (2006) find that consumers label money by income type such that income framed as a bonus—a positive departure from status quo—is more readily spent than income framed as a rebate—a return to status quo (see also Milkman and Beshears 2009 for the effects of financial windfalls on consumption). However, it is unclear whether such effects also exist in the context of point currencies. We contribute to this literature by studying how point earning characteristics affect point spending behaviors.

3. Model of Consumer's Payment Choice

Consider a consumer who wishes to purchase a product and decides whether to pay for it either with points or with money.[4] Specifically, we model consumers as choosing between paying with points or money, conditional on purchasing a product, and we do not consider consumers' product choices and the no-purchase option.[5] We present a random utility framework incorporating several

[4] Following past research on consumers' choices to spend points or money (e.g., Drèze and Nunes 2004), we assume that a consumer's budget does not constrain the use of one payment type more than the other; that is, the consumer has sufficient balances for both money and points.

[5] In Online Appendix E.3, we incorporate product choice into our model of payment currency choice with a nested logit choice model. While this nested model was not presented as our primary model due to its computational intensity and the need to simulate unobserved product choice sets, it confirms our main results and conclusions.

behavioral components, such as the subjective perceived value of points, mental accounting (to model consumers who may use separate concave cost functions to perceive the disutility of spending points and disutility of spending money), and the reference points exchange rate (to capture the perceived transaction utility from purchases where the offered exchange rate between money and points differs from a reference or benchmark exchange rate), and explicitly model the effect of consumers' points earning characteristics on their attitudes toward points.

The consumer makes payment choices by comparing the utilities of paying with money versus points. Specifically, we define consumer i's utility of paying with money or points during the decision occasion j respectively as $W^0_{ij} = U^0_{ij} + \epsilon^0_{ij}$ and $W^1_{ij} = U^1_{ij} + \epsilon^1_{ij}$, where U^0_{ij} and U^1_{ij} represent the systematic parts and ϵ^0_{ij} and ϵ^1_{ij} represent the stochastic disturbances. The consumer chooses to pay with points if $W^1_{ij} > W^0_{ij}$, which occurs with probability $q_{ij} = \mathbb{P}\left(U^1_{ij} - U^0_{ij} > \epsilon^0_{ij} - \epsilon^1_{ij}\right)$. Following the literature on discrete choice modeling (e.g., McFadden 1973, Ben-Akiva and Lerman 1985), we assume that $\epsilon^0_{ij} - \epsilon^1_{ij}$ is logistically distributed with mean 0 and scale 1. For all consumers i and decision occasions j, we can then model the observed binary payment choice y_{ij}, with $y_{ij} = 0$ representing a money payment and $y_{ij} = 1$ representing a points redemption, as follows:

$$\begin{aligned}
y_{ij} &\sim \text{Bernoulli}(q_{ij}), \\
\text{logit}(q_{ij}) &= U^1_{ij} - U^0_{ij} \\
&= -c_i \underbrace{-\Big[(p_{ij} + r_{ij})h_{ij}\Big]^{a_{p,i}}}_{\substack{\text{disutility of}\\ \text{spent/forgone points}}} \underbrace{+\Big[m_{ij}\Big]^{a_{m,i}}}_{\substack{\text{utility of}\\ \text{saved money}}} \qquad (1) \\
&\quad \underbrace{+b_{gain,i} \max\Big(\lambda_{ij} - \bar{\lambda}, 0\Big)}_{\substack{\text{transaction utility}\\ \text{from good points deal}}} \underbrace{-b_{loss,i} \max\Big(\bar{\lambda} - \lambda_{ij}, 0\Big)}_{\substack{\text{transaction disutility}\\ \text{from bad points deal}}}, \qquad (2) \\
\log(h_{ij}) &= \gamma_0 + \boldsymbol{\gamma}_i' \boldsymbol{v}_{ij} + \delta_{i,1}, \qquad (3) \\
\text{logit}(a_{p,i}) &= \beta_0 + \boldsymbol{\beta}' \boldsymbol{w}_i + \delta_{i,2}, \qquad (4)
\end{aligned}$$

where $\text{logit}(x) = \log[x/(1-x)]$ denotes the logit of a probability $x \in (0,1)$, m_{ij}, p_{ij} and r_{ij} denote the money price, points price and earned points if a money payment is made, respectively. On line (1), the parameter $c_i \in (-\infty, \infty)$ denotes an intercept term, $h_{ij} \in (0, \infty)$ denotes the subjective perceived value of points, $a_{p,i}, a_{m,i} \in (0,1)$ denote the mental accounting parameters for points and money, respectively. On line (2), the offered points exchange rate $\lambda_{ij} = m_{ij}/p_{ij}$ is compared against the reference points exchange rate $\bar{\lambda}$, and $b_{gain,i}, b_{loss,i} \in (0, \infty)$ denote reference points exchange rate sensitivities for gains and losses, respectively. On lines (3) and (4), $\boldsymbol{v}_{ij}$ and $\boldsymbol{w}_i$ are vectors of explanatory variables (capturing, e.g., points earning characteristics), the parameters γ_0 and β_0 are intercept terms, $\boldsymbol{\gamma}_i$ and $\boldsymbol{\beta}$ are vectors of slope coefficients, and both $\delta_{i,1} \sim N(0, \nu_1^2)$ and $\delta_{i,2} \sim N(0, \nu_2^2)$ are independent error terms.

Our full model captures mental accounting, reference rate effects, and consumers' points earning characteristics. However, it also nests simpler models. In our results presented in Section 6.1, we begin by testing the simplest setting in which none of these effects are present (i.e., consumers mentally account for money and points in the same way, do not consider the reference points exchange rate, and their points earning characteristics have no effect on their payment choice decisions), and incrementally increase the complexity of the model by including more behavioral components. In the following subsections, we explain our model specification in detail, starting with line (1), then describing line (2), and finishing with (3) and (4).

3.1. Subjective perceived value and mental accounting

Line (1) of our model captures the consumer's mental accounting of points versus money. Past research suggests that the perceived value assigned to points is subjective and may change even between different decision occasions (e.g., Basumallick et al. 2013). To capture such decision-level heterogeneity, we model the consumer i's subjective perceived value of a point at decision occasion j with the parameter h_{ij}. Using h_{ij}, we can convert the points price into its money equivalent price, which provides a natural way for consumers to compare the points price and money price in the same units when making a payment decision. Specifically, if consumer i chooses to redeem points at decision occasion j, they incur a loss of $p_{ij} + r_{ij}$ points from redeeming p_{ij} points and forgoing r_{ij} points that would have been rewarded had they chosen to pay with money instead. The consumer values the points loss at a money equivalent of $(p_{ij} + r_{ij})h_{ij}$. However, if the consumer chooses to pay with money, they lose m_{ij} dollars.

Next, we describe the mental accounting components. Past research suggests that consumers mentally account for money in different ways, depending, for example, on how money is earned (e.g., Thaler 1999). Similarly, for point currency, past empirical observations indicate that consumers may weigh the utility of a points redemption against the utility of a money payment in separate mental accounts (e.g., Stourm et al. 2015, Drèze and Nunes 2004). Therefore, consistent with the findings of diminishing sensitivities in losses from prospect theory (Tversky and Kahneman 1992), we model the disutility of a points loss and the disutility of a money loss separately with two concave cost functions[6] with mental accounting parameters $a_{p,i}$ and $a_{m,i}$, respectively. Specifically, the disutility or cost of paying m_{ij} dollars is $m_{ij}^{a_{m,i}}$ and that of paying p_{ij} points is $[(p_{ij} + r_{ij})h_{ij}]^{a_{p,i}}$. Our construction of these cost functions imply diminishing sensitivities, such that incremental increases in money equivalent prices result in progressively smaller increases in disutilities.

This construction also suggests that, all else being equal, a consumer with, say, $a_{m,i} > a_{p,i}$ has a money cost function that lies above and is steeper than the points cost function for money

[6] Our main results and conclusions are robust to relaxing the domain of $a_{p,i}, a_{m,i} \in (0, 1)$ to $a_{p,i}, a_{m,i} \in (0, 3)$, so that convex cost functions are possible.

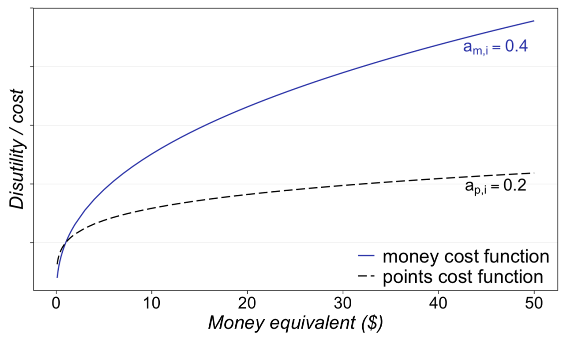


Figure 1 **Illustration of money and points cost functions when $a_{m,i} > a_{p,i}$.**

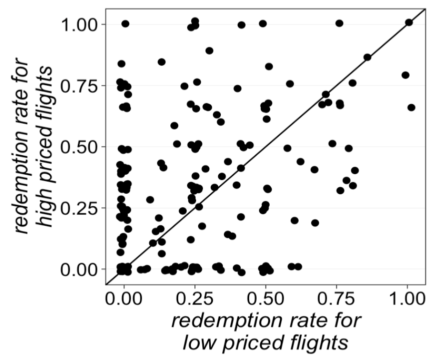
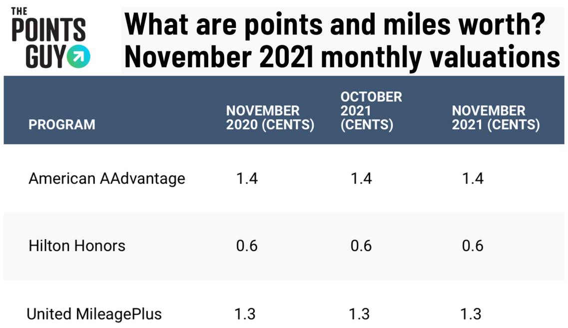


Figure 2 **Redemption rates of high versus low priced flights. High (low) priced flights are defined as flights priced above (below) the median.**

equivalent prices greater than $1. The disutility of paying with money is therefore greater than the disutility of paying with points for a given money equivalent level. In addition, as the money equivalent prices increase, the disutility of paying with money increases more than the disutility of paying with points. Therefore, higher money prices or more expensive products and services will magnify the differences between the money and points cost functions, making the points redemption more attractive relative to the money payment. These relationships are illustrated in Figure 1 that graphs the cost functions for $a_{m,i} = 0.4$ and $a_{p,i} = 0.2$.

In this example, if the consumer prefers paying with money when the money price is low, the greater increase in the disutility of money payment can result in them preferring to pay with points when the money price increases. Indeed, in model-free evidence shown in Figure 2, we observe many consumers exhibiting such reversals of preferences as prices increase (controlling for exchange rates), consistent with different mental accounting parameters for money and points. For example, consumers in the upper left of the plot have very low redemption rates when prices are low. They then prefer paying with money, as in our example. However, their preferences reverse when prices are high, since they now exhibit very high redemption rates. The opposite is true for the bottom right of the plot. Overall, the many points off the diagonal suggest that there can be high heterogeneity in mental accounting among the consumers and that the consumers are unlikely to treat points and money equally.

In addition to support from the literature and model-free evidence for the incorporation of mental accounting in our model, note also that our model nests the case in which there is no difference in consumers' mental accounting between money and points. If deemed appropriate by data, our model can estimate $a_{m,i} \approx a_{p,i}$. In Section 6.1, we also explicitly compare a model in which there are no mental accounting differences between money and points (i.e., we set $a_{m,i} = a_{p,i}$ for all i) to other models in which there are mental accounting differences. We show that the latter models exhibit better fit to the data and yield more accurate out-of-sample predictions of consumer behaviors.



THE POINTS GUY

What are points and miles worth? November 2021 monthly valuations

PROGRAM	NOVEMBER 2020 (CENTS)	OCTOBER 2021 (CENTS)	NOVEMBER 2021 (CENTS)
American AAdvantage	1.4	1.4	1.4
Hilton Honors	0.6	0.6	0.6
United MileagePlus	1.3	1.3	1.3

Figure 3 Screenshot of loyalty point valuations (in cents per point) from The Points Guy website.

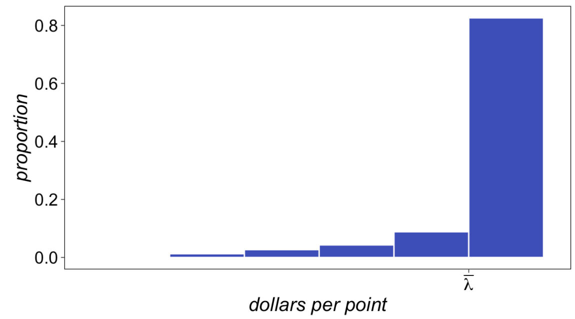


Figure 4 Histogram of points redemption exchange rates.

In summary, based on the above definitions, we define the systematic part of the utility for the money payment as $V_{ij} - [m_{ij}]^{a_{m,i}}$ and for the point payment as $V_{ij} - [(p_{ij} + r_{ij})h_{ij}]^{a_{p,i}}$, where V_{ij} represents the reservation value of the product or service purchased by the consumer. In the difference of these two utility terms the reservation value cancels, leaving us with

$$\underbrace{-\Big[(p_{ij} + r_{ij})h_{ij}\Big]^{a_{p,i}}}_{\substack{\text{disutility of} \\ \text{spent/forgone points}}} + \underbrace{\Big[m_{ij}\Big]^{a_{m,i}}}_{\substack{\text{utility of} \\ \text{saved money}}},$$

which, after adding in the intercept term $-c_i$, aligns with line (1) in our model.

3.2. Reference points exchange rate effect

Numerous websites and blogs provide advice to consumers by publishing point valuations (i.e., the typical exchange rates between points and money) for different loyalty programs. To illustrate, Figure 3 shows the point valuations for a few loyalty programs from The Points Guy website. These point valuations estimate the typical monetary value of a point and hence can help consumers to understand what they can reasonably expect to receive in exchange for their points. In particular, by comparing the reference exchange rate against the points exchange rate of a potential purchase, consumers can identify both good and bad deals. Figure 4 explores our data and finds a sharp increase in the number of point redemptions above a level $\bar{\lambda}$, offering model-free evidence of consumers' point redemptions being affected by a reference exchange rate.

Previous research on monetary prices (e.g., Thaler 1985) has modeled the total utility of a purchase as the sum of the acquisition utility (net utility of paying a store price to acquire a good) and the transaction utility (net utility of paying a store price when the expected or reference price differs from the store price). Transaction utility then captures the favorability of a deal. Such reference price effects are commonly modeled as being proportional to the difference between the store price and the reference price (Mazumdar et al. 2005). In addition, there is some evidence suggesting that reference price effects may be asymmetric such that the disutility of a bad deal is different from the utility of an equally good deal (e.g., Erdem et al. 2001).

Our model captures transaction utility that is proportional to the difference between the offered points exchange rate λ_{ij} and the reference exchange rate $\bar{\lambda}$. To allow asymmetric effects, the difference is multiplied by $b_{gain,i} > 0$ or $b_{loss,i} > 0$, depending on whether the deal is good or bad, respectively. Intuitively, $b_{gain,i}$ represents consumer i's propensity to seek good point deals and $b_{loss,i}$ represents consumer i's aversion to bad point deals. Putting this all together, the additive transaction utility becomes

$$\underbrace{b_{gain,i} \max\left(\lambda_{ij} - \bar{\lambda}, 0\right)}_{\substack{\text{transaction utility} \\ \text{from good points deal}}} \underbrace{- b_{loss,i} \max\left(\bar{\lambda} - \lambda_{ij}, 0\right)}_{\substack{\text{transaction disutility} \\ \text{from bad points deal}}},$$

which represents line (2) of our model.

Finally, given that our model nests the case in which there are no reference exchange rate effects, it can always choose to remove the exchange rate effect and set $b_{gain,i} = b_{loss,i} \approx 0$ if the data deem this appropriate. Our data, however, supports the reference exchange rate effect. In particular, in Section 6.1, we explicitly compare a model in which there are no reference exchange rate effects to other models with such effects, and we show that the latter models fit the data better and have superior out-of-sample predictive accuracies of consumer behaviors.

3.3. Points earning characteristics

To understand the effects of points earning characteristics on consumers' attitudes toward points, we explain key model parameters in terms of relevant variables such as points earning sources and other points or loyalty program related information. Equations (3) and (4) in our model capture the effects of such variables on consumers' attitudes toward points as described by the parameters h_{ij} (reflecting the subjective perceived value of a point) and $a_{p,i}$ (reflecting the disutility of paying with points). To ensure a non-negative value for the subjective perceived value h_{ij}, we model it as a log-linear function of points related decision level explanatory variables $\boldsymbol{v}_{ij}$, such as bonus point earnings and points balance at the time of the transaction (see Section 4.2 for more details):

$$\log(h_{ij}) = \gamma_0 + \boldsymbol{\gamma}_i' \boldsymbol{v}_{ij} + \delta_{i,1}.$$

Similarly, we model the mental accounting parameter $a_{p,i}$, which is always between 0 and 1, as a logit-linear function of points related consumer level explanatory variables $\boldsymbol{w}_i$, such as the amount of points earned from airline sources and elite status in the loyalty program (see Section 4.2 for more details):

$$\text{logit}(a_{p,i}) = \beta_0 + \boldsymbol{\beta}' \boldsymbol{w}_i + \delta_{i,2}.$$

Table 1 **Summary statistics for airline ticket purchases, showing the mean and 5 quantiles.**

	Mean	5%	25%	50%	75%	95%
# of purchases per member	17.7	8	9	13	20	46
# of point redemptions per member	3.8	1	2	3	5	10
# of money payments per member	13.9	3	7	10	16	40
Points redemption rate (%, member level)	25.9	5	12	21	33	67
Points price (difference from mean)	0	-9,171	-5,571	-2,087	2,804	14,484
Money price (difference from mean)	0	-119	-71	-26	39	181

4. Data and Loyalty Program Description

4.1. Data details

We obtain proprietary loyalty program transaction data from a major U.S. airline company[7]. The data describe over 29,000 unique loyalty points earning (accrual) and spending (redemption) transactions made during two recent years by 500 loyalty program consumers, who each purchased at least 8 airline tickets and made at least one money and one points payment. The threshold of 8 corresponds to relatively active consumers who purchased on average 1 one-way airline ticket each quarter or 1 round trip ticket every 6 months over the 2 year period. The ratio between the money price and points price (i.e., the points exchange rate) varies across purchases.

In this loyalty program, points do not expire and consumers can use their earned points to purchase airline tickets or other non-airline products and services. However, given that the vast majority of points redeemed (about 95%) are used as payment for airline tickets, we focus our study on consumers' redemption choices for airline tickets. Specifically, in our analysis, a decision occasion is the event when a consumer decides to use either points or money to pay for a one-way airline ticket. We define a one-way flight as one that flies from airport A to a different airport B, regardless of whether it flies direct or has a layover. In each case, the consumer makes one payment decision and the transaction shows up as a single observation in our dataset. Consumers can choose to pay entirely by either money or points but not by a combination of both. As we observe consumers' point balances at the time of each decision, we can explicitly compare their point balances and the point prices to confirm whether they can make payment choices between money and points. In total, our dataset describes about 9,000 decision occasions across 500 consumers. In about 22% of these decision occasions, the consumers pay with points. Table 1 shows some summary statistics for the data, with the actual points and money prices masked to preserve data confidentiality.

Available information about a decision occasion depend on whether the consumer paid with money or points. For an airline ticket purchased with money, our data describe the flight details at the one-way flight level, including the origin-destination route, the flight departure date, the

[7] The identity of the airline and the exact period are not revealed to preserve anonymity.

money price and earned points for the ticket purchase. For an airline ticket purchased with points (i.e., points redemption), the data describe flight details at the one-way flight level, including the origin-destination route, the redemption or booking date, the flight departure date, and the money and points prices for the redeemed ticket.

Other information in our data include consumer-level characteristics such as the consumers' program enrollment dates and whether they have elite tier statuses. In addition, each time a consumer earns points, our data describe the source and type of the point. Sources include purchases with the focal airline, spending on the airline's co-branded credit card, converting reward points from non-co-branded credit cards (or third party wallets) to the focal airline's loyalty points, and other spending with non-bank partners such as hotels and car rental companies. Point types include non-bonus (regular) points and bonus or promotional points such as those awarded for reaching a spending threshold on a co-branded credit card or promotional points for airline ticket purchases.

Although our dataset provides rich transaction information, there are some variables that are not directly recorded. For an airline ticket purchased with points, we compute r_{ij}, the points that could have been earned with a money purchase, by referencing the loyalty program's point accrual policy. Depending on the consumer's elite status and whether the booking is done online directly with the focal airline, purchasing an airline ticket with money rewards a fixed rate of points per dollar. Therefore, we can compute r_{ij} based on the observed money price m_{ij}, the consumer's elite status, and whether the consumer would have made an online booking had the consumer chosen to pay with money instead. Fortunately, each consumer's elite status at the time of the booking is observed. Our data also show whether the consumer made an online ticket redemption. If they did, we assume that the consumer would have booked online even if they had chosen to pay with money instead of points.

For an airline ticket paid with money, the booking date and point price p_{ij} are also not directly recorded, and we therefore construct proxies for these variables. To obtain a booking date for an airline ticket purchased with money, we calculate the median time between the booking and departure dates of all redemptions made by the consumer. The booking date for the ticket purchased with money is then computed by subtracting the consumer's median booking window from the observed departure date. To obtain a point price p_{ij} for an airline ticket purchased with money, we match the purchase with airline ticket redemptions (for which we observe both m_{ij} and p_{ij}) based on the origin-destination and departure date. Specifically, we find the airline ticket redemptions for the same origin-destination route and with departure dates within ± 2 weeks of the departure date of the ticket paid with money. We then compute the average points exchange rate $\hat{\lambda}$ of the matched redemptions and obtain the points price of the ticket paid with money by setting $p_{ij} = m_{ij}/\hat{\lambda}$. An average of 8.7 matched redemptions with an average coefficient of variation of 0.04 are used to proxy the points price for each ticket paid with money. This represents low dispersion relative to the mean of the exchange rates for each matched set. Similar approaches have been widely used in the literature (e.g., Kopalle et al. 2012) where prices for non-purchased brands are often not recorded and have to be matched to recorded prices paid by other consumers for the same brands.

The unavailability of booking dates and point prices p_{ij} for airline tickets paid with money represent the biggest limitation of our dataset. The proxies that we construct for these variables are treated in the same way as the observable data variables in the model, and it is therefore important for us to gauge the impact of measurement errors on the results. It is also important to assess whether our results are robust to the proxy methods that are used. In Online Appendix E, we perform extensive analyses and show that our main results are robust to these proxies. Specifically, in Online Appendix E.1, we show that our results remain robust to a different proxy method for booking dates, in which a prediction model is trained on the set of airline ticket redemptions with known booking dates. We also show that our results are robust to shifts and perturbations in the proxied booking dates, and we present explanations for this observed robustness. In Online Appendix E.2, we show that our proxy method for point prices has good predictive performance on a test set of redemption transactions with known rates and prices. We also show that our main results remain robust to different proxy methods and to perturbations of the proxied variables.

Finally, we set the reference points exchange rate $\bar{\lambda}$ in our model by grid search. Specifically, we estimate our full model using different values of $\bar{\lambda}$ and compare model fit using the criteria of Watanabe-Akaike or widely applicable information criterion (WAIC) and leave-one-out cross validation (LOO-CV) approximated within-sample. These measures are described in greater detail in Section 6.1. Both model fit criteria lead to similar conclusions, and we choose the reference exchange rate $\bar{\lambda}$ that has the best model fit and performance across both criteria.

4.2. Explanatory variables

Table 2 presents the points related explanatory variables that form the vectors $\boldsymbol{v}_{ij}$ and $\boldsymbol{w}_i$ in (3) and (4) of our model. The top panel in Table 2 shows the consumer level variables in $\boldsymbol{w}_i$. AIR_i and $CARD_i$ capture consumers' points earning source effects, specifically, the effects of earning points from airline and co-branded credit card sources. AIR_i and $CARD_i$ are not temporally associated with the decision occasions (for more information, see online Appendix A). Therefore, we include them as consumer level variables. Consumers can also earn points from other sources

[8] As a robustness check, we estimated a model with an additional dummy variable that indicates whether the ticket is a business class fare or otherwise. There is little variation in this indicator, as 98.7% of tickets are non-business class fares, and we find that including this variable does not affect the main results.

[9] The upper bound of this variable is 1 and we restrict its lower bound to -2, so that hoarding very large point balances does not have an outsized effect on this variable, since the effect of points balance is separately captured.

Table 2 **Explanatory variables in $\boldsymbol{w}_i$ and $\boldsymbol{v}_{ij}$.**

Explanatory variables in $\boldsymbol{w}_i$	
AIR_i	consumer i's standardized mean monthly airline point earnings over the data period
$CARD_i$	consumer i's standardized mean monthly co-branded card point earnings over the data period
$TOTAL.EARN_i$	consumer i's standardized mean monthly total point earnings over the data period
$TENURE_i$	standardized tenure of consumer i's program membership in years, measured from enrollment to the end of the data period
$ELITE_i$	dummy variable for consumer i's elite tier status (1: elite, 0: non-elite)
Explanatory variables in $\boldsymbol{v}_{ij}$[8]	
$BONUS_{ij}$	consumer i's standardized mean monthly bonus point earnings, up to the decision occasion j
$WALLET_{ij}$	consumer i's standardized mean monthly third party wallet point earnings, up to the decision occasion j
$EARN_{ij}$	consumer i's standardized mean monthly total point earnings, up to the decision occasion j
$SPEND_{ij}$	consumer i's standardized mean monthly total point spending, up to the decision occasion j
$DEPLETE_{ij}$	consumer i's points depletion relative to their personal targeted redemption level at decision occasion j, $\max\left(\frac{\text{target redemption level}-\text{points balance}}{\text{target redemption level}}, -2\right)$[9]
$BALANCE_{ij}$	consumer i's standardized points balance at decision occasion j

besides airline and co-branded credit cards, but these sources are not included explicitly as separate variables. Instead, they serve as the base category. We control the total level of point earnings with $TOTAL.EARN_i$ which allows us to interpret the coefficient of, say, AIR_i as the effect of increasing the level of airline earnings while holding the level of total earnings fixed or, equivalently, increasing airline earnings at the expense of earnings from other sources.

Other control variables include $TENURE_i$ and $ELITE_i$. $TENURE_i$ captures the effects of program familiarity in terms of how long consumer i has been enrolled in the loyalty program, and $ELITE_i$ captures the effects of consumer status. These variables exhibit relatively low variations across decisions and are naturally consumer level variables. Specifically, consumers' average tenure at the end of the data period is 7.8 years, but variations in tenure between decisions are only in the order of 0.1 years, given an average of 18 decision occasions per consumer over the 2 years (as presented in Table 1). Similarly, when consumers qualify for elite status, they retain this status for the current and the next calendar year, leading to stable statuses over our 2 years of data.[10]

The bottom panel in Table 2 shows the decision level variables in $\boldsymbol{v}_{ij}$. $EARN_{ij}$ controls for the level of earned points from all sources up to the decision occasion. Consumers can earn either bonus or non-bonus points. To separate between these two types, $BONUS_{ij}$ describes the bonus points earned by the consumer. These points are available only temporarily and their accruals are often followed shortly by a points redemption (see online Appendix A for more details). Therefore, we

[10] We also perform a model check in which the point mental accounting parameter is modeled as $\text{logit}(a_{p,ij}) = \beta_0 + \boldsymbol{\beta}'\boldsymbol{w}_{ij} + \delta_{i,2}$, so that the explanatory variables in $\boldsymbol{w}_{ij}$ are modeled at the decision level instead. This model performs significantly worse in estimated out-of-sample predictive accuracy.

include $BONUS_{ij}$ as a decision level variable. $WALLET_{ij}$ captures the effects of earning points from third party point wallets such as non-co-branded credit cards, as described in Section 4.1. This is included as a decision level variable because it has temporal characteristics similar to those of $BONUS_{ij}$ (see online Appendix A for more details). $SPEND_{ij}$ controls for the effects of points spending behavior before the decision occasion.

Finally, $DEPLETE_{ij}$ captures the effect of consumers being driven to earn instead of redeem points because their point balances are depleted relative to their desired or targeted redemption levels (Dorotic et al. 2014). Similarly to Dorotic et al. (2014), we measure this effect in terms of the relative distance between consumer i's points balance and their targeted redemption level, defined as the consumer's average past points redemption level. If there are no past observed redemptions, the targeted redemption level is set to the population's average past points redemption level. $BALANCE_{ij}$ captures points balance effects at decision occasion. These two variables are naturally included at the decision level because they describe the consumers' point balances at the time of making their payment choices, and the point balances fluctuate over time (there are over 29,000 points earning and spending transactions over our 2 year data period as described in Section 4.1).

5. Hierarchical Bayesian Estimation Framework

In this section, we detail a hierarchical Bayesian specification of our model. Hierarchical Bayesian models are increasingly utilized in the social sciences to model heterogeneity in consumer preferences and sensitivities as they provide a unified framework for inference and the proper accounting of parameter and model uncertainty (Rossi et al. 2005). Specifically, the hierarchical Bayesian approach provides several advantages for our analysis. First, it allows us to explicitly capture the heterogeneity across consumers, which is critical in understanding consumers' payment choice behaviors. Second, it allows us to obtain improved estimates of the consumer-level parameters by "borrowing strength" across consumers (Gelman et al. 2013). This can be particularly important for estimating parameters of consumers with a lower number of recorded decision occasions in our dataset. For instance, modeling $\text{logit}(a_{p,i})$ as a linear function of consumer level explanatory variables $\boldsymbol{w}_i$ in (4) can improve our estimates of $\text{logit}(a_{p,i})$ by borrowing strength more from consumers with similar $\boldsymbol{w}_i$ variables (George et al. 2017). Finally, the hierarchical Bayesian approach provides a joint posterior distribution of all model parameters, which allows us to assess the uncertainty of parameter estimates and any other results that depend on these estimates.

Let an n-vector $\boldsymbol{\theta}_i = [c_i, \text{logit}(a_{m,i}), \log(b_{gain,i}), \log(b_{loss,i}), \boldsymbol{\gamma}_i', \text{logit}(a_{p,i})]'$ hold the model parameters for consumer i. Our hierarchical Bayesian specification[11] then is as follows:

[11] The transaction utility variables that are associated with $b_{gain,i}$ and $b_{loss,i}$ (i.e., $\max(\lambda_{ij} - \bar{\lambda}, 0)$ and $\max(\bar{\lambda} - \lambda_{ij}, 0)$, respectively) are assessed in units of cents per point multiplied by a scaling of 10, so that $b_{gain,i}$ and $b_{loss,i}$ have the same scale as the other estimated parameters. In addition, setting the prior mean of γ_0 to zero gives h_{ij} an appropriate scale in the unit of cents per point.

$$
\begin{aligned}
&\text{for all } i \text{ and } j: && y_{ij}|\boldsymbol{\theta}_i, \gamma_0, \delta_{i,1} \sim \text{Bernoulli}(q_{ij}), \\
&\text{for all } i: && \boldsymbol{\theta}_i|\boldsymbol{\mu}, \beta_0, \boldsymbol{\beta}, \boldsymbol{\sigma}, \boldsymbol{\Omega} \sim MVN\left(\begin{bmatrix} \boldsymbol{\mu} \\ \beta_0 + \boldsymbol{\beta}'\boldsymbol{w}_i \end{bmatrix}, \text{diag}(\boldsymbol{\sigma}) \times \boldsymbol{\Omega} \times \text{diag}(\boldsymbol{\sigma})\right), \\
&\text{for } l = 1, ..., n-1: && \mu_l|\tau_{\mu_l} \sim N(0, \tau_{\mu_l}^2), \\
&\text{for } l = 0, ..., |\boldsymbol{\beta}|: && \beta_l|\tau_{\theta_l} \sim N(0, \tau_{\beta_l}^2), \\
&\text{for } l = 1, ..., n: && \sigma_l|b \sim \text{Half-Cauchy}(0, b), \\
& && \boldsymbol{\Omega}|\rho \sim \text{LKJCorr}(\rho), \\
& && \gamma_0|\tau_{\gamma_0} \sim N(0, \tau_{\gamma_0}^2), \quad \delta_{i,1}|\nu_1 \sim N(0, \nu_1^2), \\
& && \nu_1|b \sim \text{Half-Cauchy}(0, b),
\end{aligned} \tag{5}
$$

where $MVN(\boldsymbol{\alpha}, \boldsymbol{\Sigma})$ is the multivariate normal distribution with mean $\boldsymbol{\alpha}$ and covariance matrix $\boldsymbol{\Sigma}$, Half-Cauchy(a, b) is the Cauchy distribution with location a and scale b constrained to the positive values (Gelman 2006), and LKJCorr(ρ) is the LKJ correlation distribution with shape ρ. The parameter $\boldsymbol{\mu}$ is a column vector with $n-1$ entries, $\boldsymbol{\beta}$ is a column vector with $|\boldsymbol{\beta}|$ entries corresponding to the explanatory variables in $\boldsymbol{w}_i$, diag$(\boldsymbol{\sigma})$ is an $n \times n$ diagonal matrix with main diagonal entries $\sigma_1, ..., \sigma_n$, and $\boldsymbol{\Omega}$ is an $n \times n$ correlation matrix.

For parameters on the logarithmic and logit scales, we let the prior scales of their population means to be 1 and 2.5, respectively. This gives preferences to values between -5 and 5 in the logit scale, which can be considered reasonable for two reasons. First, changes outside of this range lead to very little changes in the probability scale (see, e.g., Gelman et al. 2008 for a similar justification of prior scales). Second, given that our explanatory variables $\boldsymbol{v}_{ij}$ and $\boldsymbol{w}_i$ in (3) and (4) are standardized, these prior scales capture reasonable effect sizes (for more details, see, e.g., the discussion in Gelman 2006, Section 2.3). Overall, these choices are weakly informative and allow for efficient posterior estimation without being overly influential relative to the data. The prior covariance matrix is decomposed into a scale n-vector $\boldsymbol{\sigma}$ and $n \times n$ correlation matrix $\boldsymbol{\Omega}$ because using a single prior distribution for the covariance matrix may be appropriate for scales but not for correlations, or *vice versa* (for more details, see, e.g., the discussions of covariance matrix decompositions in Gelman and Hill 2006 and Stan Development Team 2024). Finally, following McElreath (2020) and Bandiera et al. (2021), we set $b = 2.5$ and $\rho = 2$.

We estimate the model parameters jointly using a state-of-the-art approach called Hamiltonian Monte Carlo (HMC). This Markov chain simulation algorithm moves around the parameter space according to the gradient of the likelihood, which often makes it computationally more efficient than random walk algorithms such as the Metropolis algorithm. Once the Markov chain has converged, it moves around the parameter space in proportion to the actual posterior distribution. The subsequent states of the chain then form a posterior sample that can be directly used in

Bayesian inference. We assess convergence with the split-$\hat{R}$ diagnostic (e.g., Gelman et al. 2013). To check for the potential of multi-modality, we run four separate chains, each with a randomly chosen starting point. Each chain is ran for a total of 3,000 steps, of which the first 2,000 steps are discarded for warm-up and the final 1,000 steps form the posterior sample. Combining the samples from all four chains gives us a final posterior sample of 4,000 draws. Based on this sample, we can perform posterior inference and calculate summary statistics such as posterior quantiles and credible intervals.

6. Empirical Results

6.1. Comparison of model fit

Before we inspect our model for insights of consumer behavior and attitude toward loyalty points, we evaluate its potential to capture innate features of our data. Even though the goal of our paper is not to find the best "predictive" model of the consumers' decisions but to understand the consumers' payment choice behaviors, we use different estimates of out-of-sample predictive accuracy to assess how effective our modeling components (reference exchange rate effect, mental accounting, etc.) are in explaining the consumer behaviors observed in the data. To place the accuracy metrics in context and gauge the contribution of each component, we compare our model against three benchmark models that represent different simplifications of our model:

1. **Benchmark 1 (without mental accounting, reference rate effects, and points earning characteristics):** $a_{m,i} = a_{p,i}, b_{gain,i} = b_{loss,i} = 0, \boldsymbol{v}_{ij} = \boldsymbol{w}_i = \mathbf{0}$, for all i and j
2. **Benchmark 2 (with mental accounting):** $b_{gain,i} = b_{loss,i} = 0, \boldsymbol{v}_{ij} = \boldsymbol{w}_i = \mathbf{0}$, for all i and j
3. **Benchmark 3 (with mental accounting and reference rate effects):** $\boldsymbol{v}_{ij} = \boldsymbol{w}_i = \mathbf{0}$, for all i and j
4. **Full model (with mental accounting, reference rate effects, and points earning characteristics)**

Benchmark model 1 represents our model described in Section 5, under three constraints. First, setting $a_{m,i} = a_{p,i}$ implies that consumers weigh the disutility of money payment and points redemption in the same cost function, thereby exhibiting no differential mental accounting between money and points. Second, setting $b_{gain,i} = b_{loss,i} = 0$ implies that consumers do not perceive transaction utility or disutility when the offered points exchange rate differs from the reference exchange rate. Third, setting $\boldsymbol{v}_{ij} = \boldsymbol{w}_i = \mathbf{0}$ implies that the subjective perceived value h_i is modelled at the consumer level and that consumers' points earning characteristics have no explicit effect on their points attitudinal parameters h_i and $a_{p,i}$. Benchmark model 2 removes the restriction on mental accounting and Benchmark model 3 removes the restrictions on mental accounting and reference rate effects.

We first compare the model fits within-sample in terms of the Watanabe-Akaike or widely applicable information criterion (WAIC) (Watanabe 2010, Gelman et al. 2013) and leave-one-out cross-validation fit (LOO-CV) approximated with a Pareto smoothed importance sampling algorithm (Vehtari et al. 2017). These expected log point-wise posterior predictive density (elpd) measures are widely used Bayesian approaches for estimating the out-of-sample predictive accuracy of a Bayesian model based on only within-sample information, and they account for model complexity and the possibility of over-fitting. Larger elpd values indicate better performance.

Next, we compare the out-of-sample model fits with the Brier score and the area under the ROC curve. The Brier score is a strictly proper scoring rule and is the mean squared error of probabilistic forecasts, so that lower scores are better. In our Bayesian setup, we compute this score as $\frac{1}{J}\sum_{j=1}^{J}(\bar{q}_{ij} - y_{ij})^2$, where $\bar{q}_{ij}$ is the posterior mean predicted probability of points redemption for decision occasion j. The Brier score ranges between 0 and 1. Always predicting a probability of 0.5 guarantees a Brier score of 0.25. Therefore, any reasonable model should yield a Brier score below 0.25. The ROC curve is a popular method for conveying the tradeoff between the true and false positive rates, as the threshold of predicted probability $\bar{q}_{ij}$ for classifying an observation as a point redemption (outcome of $y_{ij} = 1$) is varied. We can summarize the performance of a classifier over all possible thresholds with the area under the ROC curve (Hastie et al. 2009), with larger values indicating better performance.

For the out-of-sample testing, we construct our training dataset using the first 1.5 years of data. We then compute the consumer level vector of explanatory variables $\boldsymbol{w}_i$ using only the training dataset. We include only consumers with at least one observed money payment and one observed point payment, with a minimum of 6 decisions overall.[12] Finally, we estimate our model parameters on the training set and test its predictive accuracy on the validation dataset constructed using the last 0.5 years of data. The consumers included in the validation dataset must also have been included in the training set for us to predict their decisions. With this method of construction, the training dataset consists of 376 consumers with 5,677 decision occasions and the validation dataset consists of a subset of 313 consumers with 1,487 decision occasions.

The model fits are presented in Table 3. Standard errors are shown in parentheses.[13] All four measures are consistent with one another and show improvements in performance as we move from Benchmark 1 to the Full model. The largest improvement occurs as we move from Benchmark 2 to 3 with the addition of the reference rate effects.

[12] In our main results, we set a threshold of 8 minimum decision occasions when we have 2 years of data. Here, we set a threshold of 6 (8×0.75) because we only use 1.5 out of the 2 full years of data.

[13] Standard errors for the Bayesian elpd measures are given by Vehtari et al. (2017). In a similar way, we compute standard errors for the Brier score by computing the sample standard deviation of the squared errors $(\bar{q}_{ij} - y_{ij})^2$ for the sample of J observations and dividing this by $\sqrt{J}$. Standard errors for the area under ROC curve are given by Hanley and McNeil (1982).

Table 3 **Comparison of model fits. Standard errors are shown in parentheses. WAIC and LOO-CV criteria are computed on the full dataset within-sample. Brier score and area under ROC curve measures are computed on the validation dataset out-of-sample. All scores improve from top (Benchmark 1) to bottom (Full model).**

Model	elpd_{WAIC}	Difference in elpd_{WAIC}	elpd_{LOO-CV}	Difference in elpd_{LOO-CV}	Brier score	Area under ROC curve
Benchmark 1	-4202 (52)		-4207 (52)		0.183 (0.006)	0.679 (0.016)
Benchmark 2	-4159 (52)	43 (8)	-4168 (52)	39 (8)	0.183 (0.006)	0.690 (0.016)
Benchmark 3	-3830 (54)	329 (24)	-3867 (55)	301 (25)	0.172 (0.006)	0.727 (0.016)
Full model	-3736 (54)	94 (15)	-3767 (55)	100 (15)	0.170 (0.006)	0.735 (0.016)

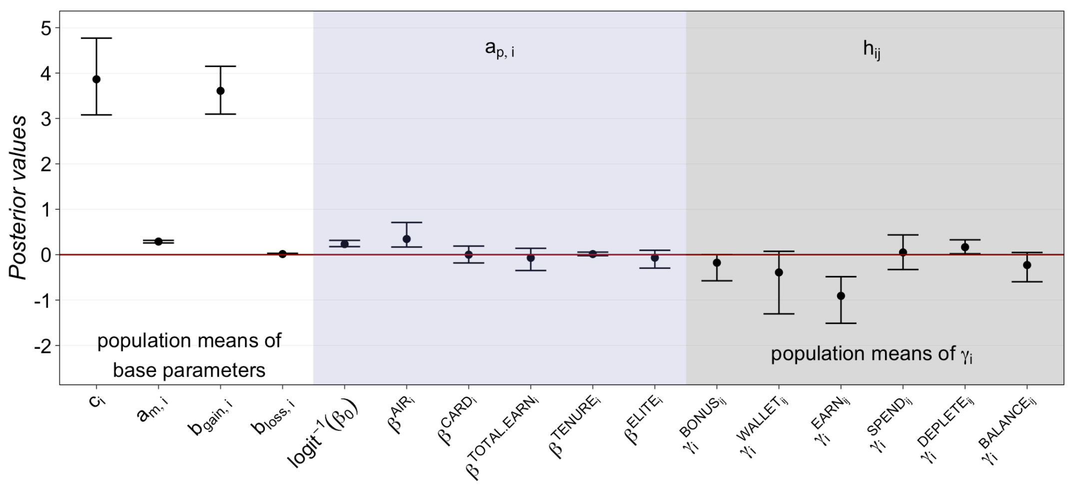


Figure 5 **The 95% credible intervals of the population means in our model (5). The blue and gray regions represent the parameters in the logit-linear model of $a_{p,i}$ and the log-linear model of h_{ij}, respectively.**

In summary, these results demonstrate the importance and advantages of including mental accounting, the reference points exchange rate effect, and points earning characteristics in a model of consumer payment choice decision. Even though here adding complexity to the model improved the model fit in each step, this does not necessarily need to be so because our performance measures take into account the trade-off between the model complexity and the added explanatory power.

6.2. Posterior estimates of parameters

Figure 5 presents the posterior estimates of the population level parameters $\boldsymbol{\mu}$ and $\boldsymbol{\beta}$ from the hierarchical Bayesian framework given in (5). The full numerical values of the posterior estimates are displayed in Online Appendix B. Recall that $\boldsymbol{\mu}$ is the mean of $(c_i, \text{logit}(a_{m,i}), \log(b_{gain,i}), \log(b_{loss,i}), \boldsymbol{\gamma}_i)$ and $\boldsymbol{\beta}$ is the vector of coefficients in the linear function of $\boldsymbol{w}_i$ that gives the mean of $\text{logit}(a_{p,i})$. These population-level parameters describe average effects of different sub-populations of consumers, and heterogeneous consumer level parameters arise from sampling distributions with these means.

The results suggest an asymmetrical reference exchange rate effect: The population means of $b_{gain,i}$ and $b_{loss,i}$ have posterior medians around 3.6 and 0, respectively. The reference points exchange rate effects are then mainly driven by the consumers' gain seeking behaviors: When faced with a good deal (i.e., $\lambda_{ij} > \bar{\lambda}$), consumers perceive significant utility boosts from redeeming points, on average, but when faced with a bad deal (i.e., $\lambda_{ij} < \bar{\lambda}$), the consumers' utilities are as if $\lambda_{ij} = \bar{\lambda}$. Therefore, from the perspective of points redemption, consumers seek good deals for points exchange rates but remain relatively insensitive to bad deals.

The literature on consumer brand choice or purchase decisions, when goods are priced only in money, models reference price effects from the money perspective (see, e.g., Mayhew and Winer 1992). Many studies find that consumers are more sensitive in the money loss region (i.e., money prices are higher than a reference level) than when they are in the money gain region. Our findings are described from the points perspective and if restated from the money perspective, they are consistent with the literature. When point exchange rates are above a reference level, consumers get more money value per point. This is a gain region from the points perspective, and the probability of points redemption q_{ij} increases by a large amount for a unit increase in point exchange rates as implied by the large $b_{gain,i}$ estimates. Restated from the money perspective, consumers get less point value per dollar of money in this same region. This implies a loss region from the money perspective, and the probability of money payment $1 - q_{ij}$ decreases by an equivalently large amount for a unit increase in point exchange rates.

Consider next the consumers' subjective perceived value of points h_{ij} that is modeled log-linearly with coefficients $\boldsymbol{\gamma}_i$. To emphasize individual elements of $\boldsymbol{\gamma}_i$, we use the superscript with the name of the explanatory variable. For instance, $\gamma_i^{BONUS_{ij}}$ denotes the coefficient of $BONUS_{ij}$.

The 95% credible intervals of the population means of $\gamma_i^{BONUS_{ij}}$ and $\gamma_i^{WALLET_{ij}}$ are largely below zero, providing evidence that higher levels of average monthly bonus or third party wallet point earnings, holding average total monthly earned points $EARN_{ij}$ and other covariates fixed, result in lower subjective perceived values h_{ij} and higher redemption propensities, on average. The 95% credible interval of the population mean of $\gamma_i^{DEPLETE_{ij}}$ is above zero, implying that an increase in points depletion (targeted redemption level − points balance) increases the subjective perceived value h_{ij} and lowers redemption propensity, on average. This indicates a tendency to defer points redemption to a later time and captures a limited degree of forward looking behavior by consumers. Next, 94.2% of the posterior mass of the population mean of $\gamma_i^{BALANCE_{ij}}$ is below zero. Therefore, with high posterior probability, higher point balances result in lower perceived values h_{ij} and higher redemption propensities. Finally, the 95% credible interval of the population mean of $\gamma_i^{SPEND_{ij}}$ is relatively wide and symmetric around zero. Therefore, we cannot make conclusions, with reasonable posterior probability, about its magnitude or direction.

Next, we consider consumers' mental accounting parameter $a_{p,i}$ which is modeled as a logit-linear function with coefficients $\boldsymbol{\beta}$. The posterior interval of β^{TENURE_i} is approximately symmetric around

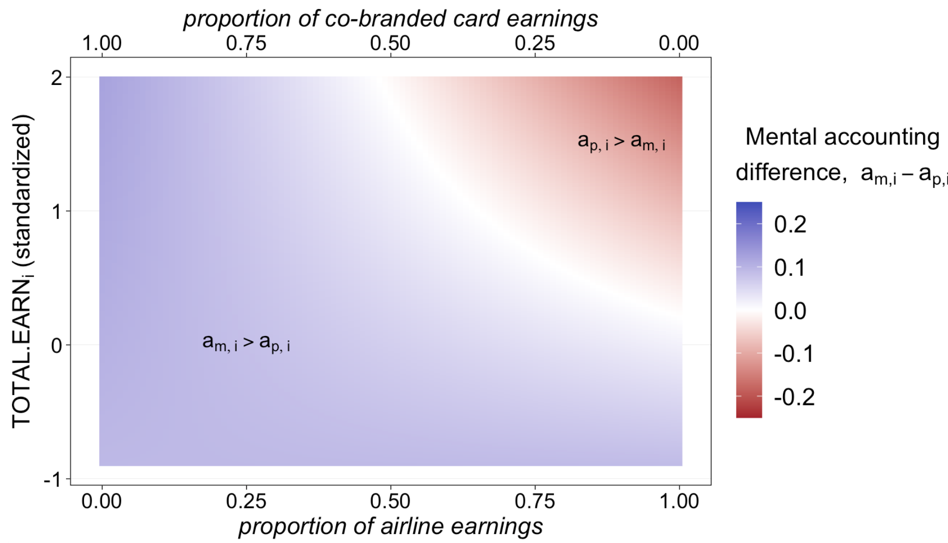


Figure 6 **Effects of point earning source variations on the mean of the mental accounting differences**

zero. Therefore we do not have strong statistical evidence to distinguish between a positive or negative association between consumers' familiarity with the loyalty program and their points mental accounting function. The posterior interval of β^{ELITE_i} also includes zero. This time, however, the median estimate is negative and 79% of the posterior mass is below zero, which we interpret as mild evidence of a negative association between elite status and $a_{p,i}$ values.

Finally, we jointly assess the estimates of β^{AIR_i}, β^{CARD_i} and $\beta^{TOTAL.EARN_i}$ and their associations with consumers' mental accounting. The quantity $a_{m,i} - a_{p,i}$ describes how consumer i perceives money and points relative to each other. If this difference is around zero, the consumer's cost functions for money and points are similar. The consumer then treats points as if they are money. But if the difference is large, the consumer accounts for money and points under different cost functions and hence treats them differently.

To illustrate how this mental accounting is expected to change with the amount and sources of the earned points, consider a consumer with average tenure ($TENURE_i = 0$) and non-elite tier status ($ELITE_i = 0$). For simplicity, we assume that there are only two sources of points, namely airline and co-branded card earnings.[14] Under different levels of total points earning and proportions of airline earnings, we can construct $\boldsymbol{w}_i$ and estimate the population mean of $a_{m,i} - a_{p,i}$. This population mean represents the expected mental accounting difference of consumers with characteristics $\boldsymbol{w}_i$. These estimates are presented in Figure 6.

Figure 6 shows that the primary point earning source and the total earning level are jointly associated with consumers' attitudes toward points and money. First, focus on the top-right corner

[14] This assumption is considered reasonable because airline and co-branded card earnings account for about 95% of the total earnings in our data.

of the plot. This region represents consumers who earn many points and mainly from spending with the airline. Holding all else fixed, such consumers' points mental accounts tend to be steeper than their money mental accounts (i.e., $a_{p,i} > a_{m,i}$). This suggests that consumers who earn many more points by flying than by using a co-branded credit card tend to value points more than money.

Next, shifting away from the top-right corner of the plot, as consumers earn less overall or more from co-branded credit cards, they begin to treat points more like money and eventually $a_{m,i} \approx a_{p,i}$. Equal mental accounting of points and money occurs at the white band shown in Figure 6. For example, consumers who earn many points from diverse sources or consumers who earn an average number of points mainly from the airline source tend to have little to no mental accounting differences between points and money.

Shifting even further away from the top-right corner of the plot and crossing the white band leads to a region where the consumers treat money as more valuable than points. In particular, consumers who earn relatively few points (i.e., $TOTAL.EARN_i < 0$) tend to exhibit higher utilities for money than points regardless of the earning source. In addition, regardless of the total points earning level, consumers who earn points mainly through co-branded credit cards tend to have lower utilities for points compared to money.

6.3. Consumer heterogeneity and policy targeting

Our hierarchical Bayesian model describes each consumer's attitude toward points with a set of idiosyncratic parameters. In total, this involves 500 sets of parameters. To draw broader insights from this population of behavioral parameters and better understand consumer heterogeneity, we propose a probabilistic segmentation of consumers. Also, through counterfactual analysis, we demonstrate how a firm can implement policies to target and influence consumers' payment choices.

For segmentation, we use the k-medoids clustering algorithm (a.k.a., Partitioning Around Medoids) that is an unsupervised learning algorithm for partitioning the observations into k clusters. Unlike the well-known k-means clustering algorithm that can choose cluster centers freely, the k-medoids algorithm must choose the cluster centers strictly among the observations, making it more robust against outliers (Hastie et al. 2009). Specifically, we segment consumers into $k = 4$ clusters based on $a_{m,i} - a_{p,i}, b_{gain,i}$ and $h_{median,i}$, where $h_{median,i}$ represents consumer i's median subjective perceived value h_{ij} across all their decision occasions j.[15] We choose $k = 4$ because this leads to distinctive segments and meaningful managerial insights (see online Appendix C for a detailed discussion). We propagate the posterior uncertainty about the consumer level parameters

[15] These segmentation dimensions include all consumer level parameters in (1) and (2), except c_i and $b_{loss,i}$. We do not segment based on c_i because it represents all unmodeled consumer level effects on redemption propensity and hence has only limited interpretability. We also exclude $b_{loss,i}$ because of its small effect size and low heterogeneity among consumers (as shown in Figure 5).

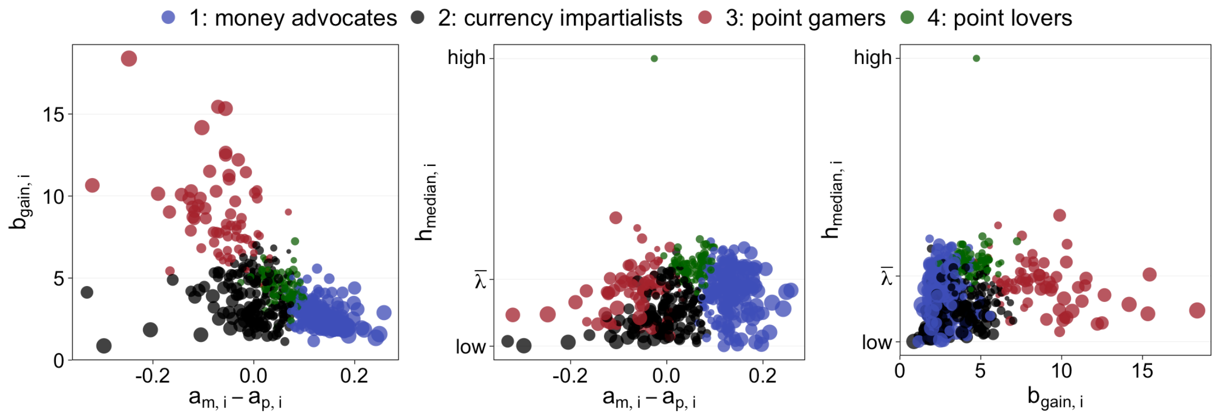


Figure 7 k**-medoids probabilistic segmentation of consumers. The points show the consumers' median posterior parameter values. The color of the points indicates the segment that the consumer is most likely to belong, and the point size is proportional to the probability of that segment membership.**

to the segmentation by applying the segmentation separately to each draw in our posterior sample. The ensemble of the draw-specific assignments then represent a probabilistic segmentation of the consumers.[16]

Figure 7 visualizes the most likely consumer segmentation with three plots. The points represent individual consumers. The color of a point describes the segment that this consumer is most likely to belong, and the size of the point is proportional to the probability of the consumer belonging to this segment. The uncovered segments have distinctive behavioral characteristics. Segment 1, in blue, consists of 191 consumers who have been labeled as "*money advocates*" because they have higher utilities for money and are more sensitive to money losses. Segment 2, in black, consists of 153 consumers who have little to no mental accounting differences and hence labeled as "*currency impartialists*". Both Segments 1 and 2 are characterized by low $b_{gain,i}$ values and hence their consumers are less likely to seek good point deals. In contrast, Segment 3, in red, is made up of 73 "*point gamers*", characterized by their high $b_{gain,i}$ values that drive them to seek good point deals. They also exhibit higher utilities for points and are more sensitive to point losses. Finally, Segment 4, in green, consists of 83 "*point lovers*" with high subjective perceived value of points $h_{median,i}$ but little to no mental accounting differences.

The left panel of Figure 8 shows consumers' points redemption rates (i.e., the proportion of decision occasions when they chose to pay with points) for each of the four segments. Segments 2 and 3 exhibit similar points redemption rates but, as was shown in Figure 7, have different behavioral drivers behind their redemption decisions: While consumers of Segment 2 tend to treat

[16] Even though this resembles the bootstrap aggregation clustering technique in Dudoit and Fridlyand (2003), to the best of our knowledge, no previous work has introduced or performed probabilistic clustering by aggregating the ensemble of segmentation results that are obtained for each draw of the posterior sample. More details of our probabilistic segmentation approach can be found in online Appendix C.

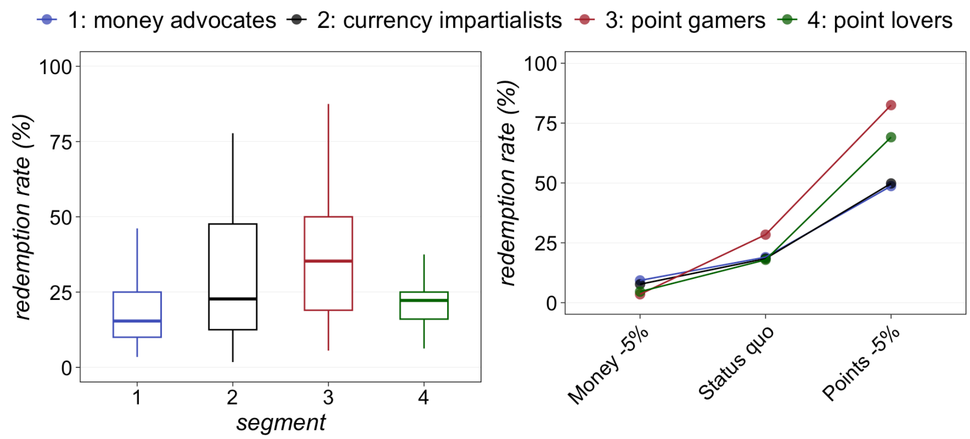


Figure 8 **Boxplots of consumers' observed point redemption rates (left), and simulated changes in redemption rates due to discounting the money price or the points price by 5% (right).**

points like money and hence choose more impartially between the two currencies, consumers of Segment 3 are mainly driven by their desire to obtain good point deals and hence redeem when points exchange rates are favorable. Similarly, Segments 1 and 4 exhibit similar points redemption rates, but consumers of Segment 1 are likely to redeem when the difference in disutilities of using points versus money is pronounced, while consumers of Segment 4 are more impartial between points and money losses but may refrain from redemption due to a high valuation of points.

This illustrates that designing targeted loyalty policies solely based on the observed redemption rates can neglect important behavioral drivers. Now, consider how discounting the money or points price affects redemption rates among the four segments. The effects of such policies can be studied with counterfactual policy simulations under our estimated model. In practice, such targeted policies could be implemented by providing discount codes to specific consumers or segments only. Using the posterior medians of the model parameters, the right panel of Figure 8 demonstrates the effects of discounting either the money price or the points price by 5% for each of the four segments. Given that the points exchange rate is $\lambda_{ij} = m_{ij}/p_{ij}$, discounting the money price also lowers the points exchange rate, whereas discounting the points price raises the points exchange rate. Therefore, as expected, *"point gamers"* with large $b_{gain,i}$ exhibit the highest sensitivities to these policies. More importantly, even though Segments 1 and 4 and Segments 2 and 3 appear similar based on observed redemption rates, in each pair one segment reacts more strongly to discounted point prices. As demonstrated, *"point lovers"* and *"point gamers"* are more sensitive to the point price discount than *"money advocates"* or *"currency impartialists"*. This clearly demonstrates why understanding heterogeneity in consumers' attitudes toward point currency is important. Without recognizing different behavioral factors, it is difficult to design optimal targeting policies.

Next, we describe the revenue impacts of price discount policies. Discounts on the money or point price incentivize consumers to pay with money or points, respectively. These can be employed during time periods when the firm assesses that either money or point payment is relatively more important,[17] and can complement the firm's existing pricing policies.

We begin by describing the direct revenue impact of a money or points payment. In accordance with the accounting standard of IFRS 15 (KPMG 2022), when the firm sells an airline ticket for m_{ij} dollars, the firm only recognizes a portion of this as revenue and defers the rest. This is because the firm also awards r_{ij} points to the consumer, which represents a performance obligation for the firm. The consumer is able to redeem the r_{ij} points in future for other products or services, and these points are monetarily valued at $r_{ij}\lambda_{mean}$ dollars on average.[18] The firm then recognizes a fraction $\eta_m^{direct} = \frac{m_{ij}}{m_{ij}+r_{ij}\lambda_{mean}}$ of the money payment m_{ij} as revenue and defers the remainder. Intuitively, the consumer has paid m_{ij} dollars in exchange for goods worth a total of $m_{ij} + r_{ij}\lambda_{mean}$, and the firm recognizes partial revenue proportional to the total value of goods.

At some point in the future, the consumer may wish to redeem the awarded points $r_{ij'}$, together with their other accrued points, for an airline ticket priced at p_{ij}. In this part, we will denote the awarded points and prices observed by the consumer during a past decision occasion with the subscript j'. In particular, the consumer paid $m_{ij'}$ dollars and was awarded $r_{ij'}$ points in the past, and is now deciding to pay with points on a present decision occasion with point price p_{ij}. When the consumer chooses to pay with points for the airline ticket and in the process redeems the past awarded points $r_{ij'}$, the firm recognizes the remainder fraction $1 - \frac{m_{ij'}}{m_{ij'}+r_{ij'}\lambda_{mean}}$ of the past money payment $m_{ij'}$ as revenue. Therefore, for each point redeemed, the firm recognizes revenue of $\eta_p^{direct} = \left(1 - \frac{m_{ij'}}{m_{ij'}+r_{ij'}\lambda_{mean}}\right)\frac{m_{ij'}}{r_{ij'}}$.[19] Taken together, the direct revenue impact of a flight purchase is $\eta_m^{direct} \cdot m_{ij} \cdot \mathbf{1}_{y_{ij}=0} + \eta_p^{direct} \cdot p_{ij} \cdot \mathbf{1}_{y_{ij}=1}$. In expectation, this is $\eta_m^{direct} \cdot m_{ij} \cdot (1 - q_{ij}) + \eta_p^{direct} \cdot p_{ij} \cdot q_{ij}$.

In addition, there are also indirect revenue impacts of money and points payment. When m_{ij} dollars are paid by a consumer, the firm receives m_{ij} dollars in incoming cash flow. The importance of this depends on the firm's cash flow and liquidity health, and may vary over time. Similarly, when p_{ij} points are paid by a consumer, the firm reduces its deferred revenue liabilities on its balance sheet, and may also improve customer engagement and loyalty. For example, through the rewarded behavior effect, consumers may increase their future purchases with the firm for each dollar value of product or service that they obtain through points redemption (Bijmolt et al.

[17] This is motivated by recent examples where some airlines offered discount codes for point redemptions and money payments. For instance, Southwest Airlines offered a 20% discount code for point redemptions in March 2023 that was valid over a 3 day booking period. Similarly, Swiss International Air Lines offered a HKD500 discount code for money payments in November 2022 that was valid over a 15 day booking period.

[18] We compute λ_{mean} in our dataset by taking the mean of all offered exchange rates for airline tickets.

[19] To illustrate using a numerical example, suppose a consumer pays \$90 for a good and is awarded 5 points, with each point redeemable for \$2 worth of goods. Given that $\eta_m^{direct} = 0.9$, the firm recognizes partial revenue of \$81 at the time of the money purchase. It defers the remainder of the revenue. Eventually, when the points are redeemed in future, the firm recognizes the deferred revenue of $\eta_p^{direct} = 1.8$ dollars per point redeemed.

2010). The importance of this depends on the firm's liability and customer engagement levels, and may also vary over time. We model such indirect expected revenue impacts parsimoniously as $\eta_m^{indirect} \cdot m_{ij} \cdot (1-q_{ij}) + \eta_p^{indirect} \cdot \lambda_{mean} p_{ij} \cdot q_{ij}$, where $\eta_m^{indirect}, \eta_p^{indirect} \geq 0$ capture the importance of the indirect revenue impacts of money and points payment, respectively, and $\lambda_{mean} p_{ij}$ represents the average money value that a consumer obtains by redeeming p_{ij} points.

The total expected revenue impact, $\mathrm{E}(\pi)$, is therefore $\mathrm{E}(\pi) = (\eta_m^{direct} + \eta_m^{indirect}) \cdot m_{ij} \cdot (1-q_{ij}) + \left(\eta_p^{direct} + \eta_p^{indirect} \lambda_{mean}\right) \cdot p_{ij} \cdot q_{ij}$. The coefficients η_m^{direct} and η_p^{direct} representing the direct revenue impacts of money and points payment, respectively, can be computed directly by the firm from data. The firm can then compute the total revenue impact of points and money discount policies given their assessments of the indirect revenue impact of points ($\eta_p^{indirect}$) and money ($\eta_m^{indirect}$). As an illustration, when $\eta_m^{indirect}$ is sufficiently high relative to $\eta_p^{indirect}$, i.e., money payments are more important, the firm may wish to incentivize money payments by providing money price discounts. This reduces the probability of points redemption q_{ij} and the expected revenue from points payment in the second part of $\mathrm{E}(\pi)$. This decreases the money price m_{ij} and increases the probability of money payment $1 - q_{ij}$. If the increase in probability of money payment is large enough, the expected revenue from money payment in the first part of $\mathrm{E}(\pi)$ increases. In this way, discounting the money price may increase expected revenues from the status quo of no discounts.

In Figure 9, we present the revenue increase of points and money discount policies from the baseline of no discounts for a range of discount levels.[20] There are two main insights here. First, when the firm implements a point discount policy to incentivize point payment, it increases its revenue from the baseline of no discounts by a larger factor than when it implements a money discount policy to incentivize money payment. This is because consumers in the program are more sensitive to point discounts than to money discounts. Second, different consumer segments have different optimal levels of point discounts, but similar optimal levels of money discounts. In the example shown in Figure 9, if the firm wishes to incentivize point payment by providing point discounts to all consumers, it can maximize revenues by providing Segments 3 and 4 of point gamers and point lovers with around 8% discounts and the remaining segments with around 12% discounts. On the other hand, if the firm wishes to incentivize money payment by providing money discounts, it can maximize revenues by simply providing all consumers with around 3% discounts.

7. Conclusion

In this study, we investigate how consumers choose to pay either with money or with points and how consumers' point earning sources are associated with the way they treat points. Our results

[20] We perform the same analysis for different levels of $\eta_m^{indirect}$ and $\eta_p^{indirect}$ in Online Appendix D. The specific levels of optimal discounts and revenue increases vary with the firm's assessed levels of $\eta_m^{indirect}$ and $\eta_p^{indirect}$ but our qualitative conclusions remain robust and unchanged.

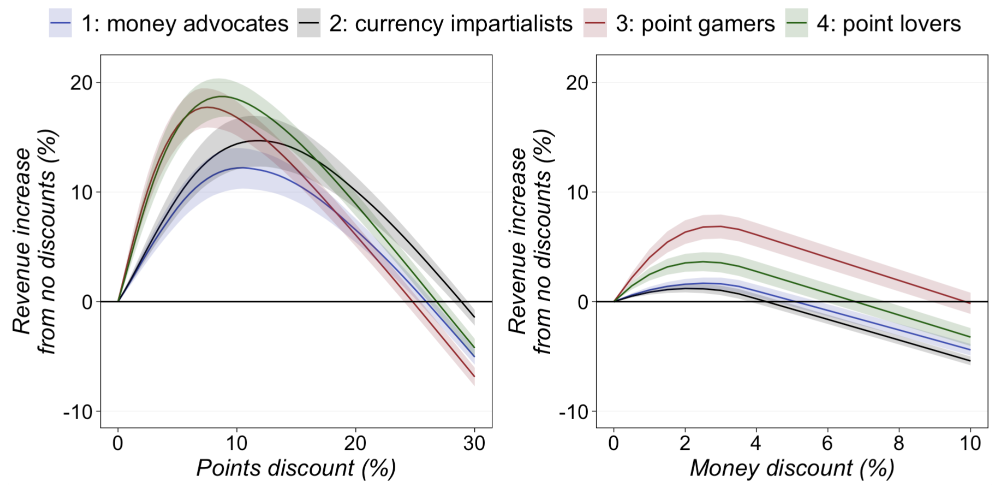


Figure 9 **Percentage revenue increase from the implementation of points and money discount policies, compared to the baseline of no discounts. Shaded regions represent 95% credible intervals. (Left) For a points discount policy, $\eta_m^{indirect} = 0, \eta_p^{indirect} = 0.5$. (Right) For a money discount policy, $\eta_m^{indirect} = 0.5, \eta_p^{indirect} = 0$.**

provide evidence that the mental accounting of money and points, the subjective perceived value of points, and the reference points exchange rate all play critical roles in influencing consumers' payment choices. We also find that consumers' primary source of point earnings and total earning levels are jointly associated with the way they mentally account for money and points. For example, consumers who earn a lot of points and mainly with the airline value points more than money but those who earn fewer points or mostly from co-branded cards value money more than points.

The effort justification effect may partially explain these findings. Past literature shows that people value an object more if they expend more effort in obtaining it. For example, Norton et al. (2012) find that participants value a self-assembled object more than an identical pre-assembled one. For many, it may take more effort to earn points from flying than by spending on co-branded credit cards, since flying is generally an infrequent activity whereas there are many opportunities to earn points from co-branded credit cards during everyday shopping activities. This effort justification effect therefore predicts that consumers value points more if they earn points mainly from flying with the airline. The results are partially consistent with this prediction.

This also indicates that the monetization of points through partnerships may have a non-trivial impact. Partnerships may increase the total earning level and decrease the proportion of earning with the focal firm. Thus, such monetization may encourage consumers to value points more, but if earnings from partnerships become dominant, consumers may begin to view points as less valuable.

To better understand heterogeneity in consumers' attitudes toward points, we propose a probabilistic segmentation and illustrate how our model and results can help firms to design more efficient targeting policies. Importantly, firms may observe similar redemption rates among consumers whose behaviors may be driven by distinctly different factors: Two different consumers may be enticed to redeem points, but one of them may be attracted by the high points exchange rate while the other may perceive low utilities for points and hence wants to spend them regardless of the exchange rate. In particular, through counter factual analysis of different point and money price discounting policies, we demonstrate how understanding heterogeneity in consumers' attitudes toward points can help the firm to design optimal targeting policies.

Although this research reveals several important factors affecting consumers' attitudes toward point currency and their decision to spend it, it is not without limitations. For example, the results on the effect of the primary point earning source on consumers' attitudes represent associations but not necessarily causations. Future research can thus expand on this study and conduct thoughtfully designed controlled experiments to learn more about the mechanism behind the effect of earning source diversification and monetization of loyalty points. Another limitation of our research is that booking dates and point prices are not available in our dataset for those airline tickets that are paid with money. We construct proxies for these variables and show in Online Appendices E.1 and E.2 that our main results remain robust to perturbations and shifts in these proxied variables and to different proxy methods. While these analyses give us greater confidence in the robustness of our results, they cannot substitute for data that is actually observed. We hope that our research will motivate future studies in this domain that may be accompanied by more complete datasets.

References

Altman D, Yom-Tov GB, Olivares M, Ashtar S, Rafaeli A (2021) Do customer emotions affect agent speed? An empirical study of emotional load in online customer contact centers. *Manufacturing & Service Operations Management* 23(4):854–875.

American Airlines (2015, 2019) American Airlines Annual Reports URL https://americanairlines.gcs-web.com/financial-results/financial-aal.

Arslan HA, Easley RF, Wang R, Yılmaz Ö (2021) Data-driven sports ticket pricing for multiple sales channels with heterogeneous customers. *Manufacturing & Service Operations Management* .

Bandiera O, Fischer G, Prat A, Ytsma E (2021) Do women respond less to performance pay? Building evidence from multiple experiments. *American Economic Review: Insights (forthcoming)* .

Baron O, Deng C, He S, Yuan H (2020) Scalable data-driven forecasting and revenue management with asymmetric formation of reference prices. *Working paper* .

Basumallick D, Ozdaryal B, Madamba-Brown C (2013) Perceived value of a mile. *Journal of Revenue and Pricing Management* 12:8–15.

Ben-Akiva M, Lerman SR (1985) *Discrete Choice Analysis: Theory and application to travel demand* (The MIT Press).

Bijmolt THA, Dorotic M, Verhoef PC (2010) Loyalty programs: generalizations on their adoption, effectiveness and design. *Foundations and Trends in Marketing* 5(4):197–258.

Breugelmans E, Bijmolt THA, Zhang J, Basso LJ, Dorotic M, Kopalle P, Minnema A, Mijnlieff WJ, Wünderlich NV (2015) Advancing research on loyalty programs: A future research agenda. *Marketing Letters* 26:127–139.

Buell RW (2021) Last-place aversion in queues. *Management Science* 67(3):1430–1452.

Carluccio J, Eizenman O, Rothschild P (2021) Next in loyalty: Eight levers to turn customers into fans. *McKinsey & Company* URL https://www.mckinsey.com/business-functions/marketing-and-sales/our-insights/next-in-loyalty-eight-levers-to-turn-customers-into-fans.

Chun SY, de Boer E (2021) How loyalty programs are saving airlines. *Harvard Business Review* .

Chun SY, Hamilton R (2021) Should I pay with money or redeem points for this purchase? Effects of exchange rate stability on loyalty point redemption. *Working paper* .

Chun SY, Iancu DA, Trichakis N (2020) Loyalty program liabilities and point values. *Manufacturing & Service Operations Management* 22(2):257–272.

Chung K (2020) Incorporating a "better" behavioral bias for both consumers and firms in rebate programs. *Management Science* 66(4):1627–1646.

Delta Air Lines (2020) Delta Air Lines Form 10-K URL https://ir.delta.com/financials/sec-filings/sec-filings-details/default.aspx?FilingId=14706066.

Dorotic M, Verhoef PC, Fok D, Bijmolt THA (2014) Reward redemption effects in a loyalty program when customers choose how much and when to redeem. *International Journal of Research in Marketing* 31(4):339–355.

Drèze X, Nunes JC (2004) Using combined-currency prices to lower consumers' perceived cost. *Journal of Marketing Research* 41(1):59–72.

Dudoit S, Fridlyand J (2003) Bagging to improve the accuracy of a clustering procedure. *Bioinformatics* 19(9):1090–1099.

Economist (2005) Frequent flyer miles: In terminal decline. *The Economist* URL https://www.economist.com/leaders/2005/01/06/in-terminal-decline.

Epley N, Mak D, Idson LC (2006) Bonus of rebate?: The impact of income framing on spending and saving. *Journal of Behavioral Decision Making* 19(3):213–227.

Erdem T, Mayhew G, Sun B (2001) Understanding reference-price shoppers: A within- and cross-category analysis. *Journal of Marketing Research* 38(4):445–457.

Ertekin N, Ketzenberg ME, Heim GR (2020) Assessing impacts of store and salesperson dimensions of retail service quality on consumer returns. *Production and Operations Management* 29(5):1232–1255.

Gelman A (2006) Prior distributions for variance parameters in hierarchical models. *Bayesian Analysis* 1(3):515–533.

Gelman A, Carlin JB, Stern HS, Dunson DB, Vehtari A, Rubin DB (2013) *Bayesian Data Analysis* (Chapman and Hall/CRC).

Gelman A, Hill J (2006) *Data analysis using regression and multilevel/hierarchical models* (Cambridge University Press).

Gelman A, Jakulin A, Pittau MG, Su YS (2008) A weakly informative default prior distribution for logistic and other regression models. *The Annals of Applied Statistics* 2(4):1360–1383.

George EI, Ročková V, Rosenbaum PR, Satopää VA, Silber JH (2017) Mortality rate estimation and standardization for public reporting: Medicare's Hospital Compare. *Journal of the American Statistical Association* 112(519):933–947.

Hanley JA, McNeil BJ (1982) The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology* 143(1):29–36.

Hastie T, Tibshirani R, Friedman J (2009) *The elements of statistical learning* (Springer, New York, NY).

Hsu T (2019) Why rewards for loyal spenders are 'a honey pot for hackers'. *The New York Times* URL https://www.nytimes.com/2019/05/11/business/rewards-loyalty-program-fraud-security.html.

Huckman RS, Staats BR, Upton DM (2009) Team familiarity, role experience, and performance: Evidence from indian software services. *Management Science* 55(1):85–100.

Keppler SM, Li J, Wu D (2021) Stopping the revolving door: An empirical and textual study of crowdfunding and teacher turnover. *Working paper* .

Kopalle PK, Sun Y, Neslin SA, Sun B, Swaminathan V (2012) The joint sales impact of frequency reward and customer tier components of loyalty programs. *Marketing Science* 31(2):216–235.

KPMG (2022) *IFRS 15 Handbook* (KPMG).

Lu Y, Musalem A, Olivares M, Schilkrut A (2013) Measuring the effect of queues on customer purchases. *Management Science* 59(8):1743–1763.

Martínez-de-Albéniz V, Sosa M, Carrera C (2020) The bright side of bad choices: Evidence from restaurant exploration. *Working Paper* .

Mayhew GE, Winer RS (1992) An empirical analysis of internal and external reference prices using scanner data. *Journal of Consumer Research* 19(1):62–70.

Mazumdar T, Raj SP, Sinha I (2005) Reference price research: Review and propositions. *Journal of Marketing* 69(4):84–102.

McElreath R (2020) *Statistical rethinking: A Bayesian course with examples in R and Stan* (Chapman and Hall/CRC).

McFadden D (1973) Conditional logit analysis of qualitative choice behavior. *Frontiers of Econometrics* P. Zarembka (ed.), Academic Press: New York:105–142.

Milkman KL, Beshears J (2009) Mental accounting and small windfalls: Evidence from an online grocer. *Journal of Economic Behavior & Organization* 71(2):384–394.

Norton MI, Mochon D, Ariely D (2012) The IKEA effect: When labor leads to love. *Journal of Consumer Psychology* 22(3):453–460.

Orhun AY, Guo T, Hagemann A (2021) Reaching for gold: Frequent-flyer status incentives and moral hazard. *Marketing Science (forthcoming)* .

Özer O, Sul I, Şimşek AS (2020) Damned if you buy, damned if you wait: An empirical investigation of customer regret under markdown pricing and its implications to retailing. *Working Paper* .

Rossi PE, Allenby GM, McCulloch R (2005) *Bayesian Statistics and Marketing* (John Wiley & Sons, Ltd).

Schlangenstein M, Bachman J (2019) Airlines want you to burn your miles. *Bloomberg* URL https://www.bloomberg.com/news/articles/2019-03-05/airlines-want-you-to-burn-your-miles.

Soman D (2001) Effects of payment mechanism on spending behavior: The role of rehearsal and immediacy of payments. *Journal of Consumer Research* 27(4):460–474.

Southwest (2010, 2019) Southwest Airlines Annual Reports URL https://www.southwestairlinesinvestorrelations.com/financials/company-reports/annual-reports.

Stan Development Team (2024) Stan modeling language users guide and reference manual, 2.34. URL https://mc-stan.org/users/documentation/.

Stourm V, Bradlow ET, Fader PS (2015) Stockpiling points in linear loyalty programs. *Journal of Marketing Research* 52(2):253–267.

Tan TF, Netessine S (2019) When you work with a superman, will you also fly? An empirical study of the impact of coworkers on performance. *Management Science* 65(8):3495–3517.

Tereyağoğlu N, Fader PS, Veeraraghavan S (2018) Multiattribute loss aversion and reference dependence: Evidence from the performing arts industry. *Management Science* 64(1):421–436.

Thaler RH (1985) Mental accounting and consumer choice. *Marketing Science* 4(3):199–214.

Thaler RH (1999) Mental accounting matters. *Journal of Behavioral Decision Making* 12(3):183–206.

Tversky A, Kahneman D (1992) Advances in prospect theory: Cumulative representation of uncertainty. *Journal of Risk and Uncertainty* 5:297–323.

Vehtari A, Gelman A, Gabry J (2017) Practical Bayesian model evaluation using leave-one-out cross-validation and WAIC. *Statistics and Computing* 27:1413–1432.

Watanabe S (2010) Asymptotic equivalence of Bayes cross validation and widely applicable information criterion in singular learning theory. *Journal of Machine Learning Research* 11(116):3571–3594.

Yu Q, Allon G, Bassamboo A (2017) How do delay announcements shape customer behavior? An empirical study. *Management Science* 63(1):1–20.

Online Appendices

Appendix A: Temporal nature of points earning source and type variables

The explanatory variables of $BONUS_{ij}$ and $WALLET_{ij}$ are modeled at the decision level due to the temporal nature of bonus points accruals and third party wallet accruals. The explanatory variables of AIR_i and $CARD_i$ are modeled at the consumer level due to the lack of such a timing component. We construct a simple ratio measure to illustrate this temporal nature. We first compute the average days between redemptions for each consumer, where we restrict the analysis to consumers who redeemed on at least 3 different dates, so that there are at least 2 time intervals between redemption dates[1]. We then compute, for each point accrual occasion, the number of days to the next redemption. The ratio measure is simply the number of days to the next redemption divided by the average days between redemptions for the consumer. If the occurrences of the accrual and redemption events are random, we expect that, on average, the ratio measure is 0.5 such that accrual events tend to occur at the midpoints of the redemption intervals. If there is a temporal component or timing associated with an accrual, we will expect that the ratio measure is smaller than 0.5 such that accrual events occur nearer to the next redemption. We plot the ratio measures for all the variables in Figure 1. Both $WALLET_{ij}$ and $BONUS_{ij}$ have median ratio measures that are below 0.5, indicating a temporal or timing component, while the medians of AIR_i and $CARD_i$ are around 0.5, indicating little to no temporal components.

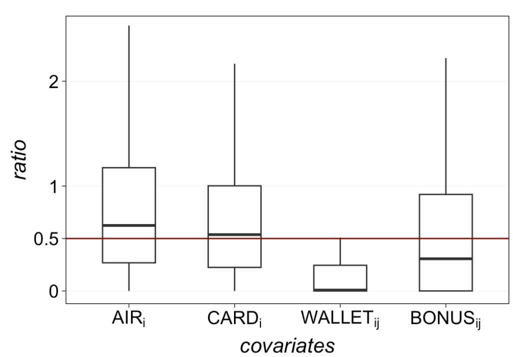


Figure 1 Boxplots of the ratio (days to next redemption) / (average days between redemptions) for points earning source and type variables. $WALLET_{ij}$ **and** $BONUS_{ij}$ **have temporal components.**

[1] In addition, for each consumer, we include 1) the interval between the start of the data period (or the consumer's program enrollment date, whichever comes later) and the first redemption, and 2) the interval between the last redemption and the end of the data period because these are also intervals without redemptions.

Appendix B: Posterior estimates of parameters

Table 1 Posterior quantiles of the population means $\boldsymbol{\mu}$ and $\boldsymbol{\beta}$.

	Parameters	2.5%	5%	50%	95%	97.5%
Population means of	c_i	3.08	3.22	3.86	4.64	4.77
base parameters	$a_{m,i}$	0.26	0.26	0.29	0.31	0.32
$(\mu_1 - \mu_4)$	$b_{gain,i}$	3.10	3.16	3.61	4.06	4.15
	$b_{loss,i}$	0.00	0.01	0.01	0.03	0.03
$\boldsymbol{\beta}$	$\text{logit}^{-1}(\beta_0)$	0.18	0.18	0.23	0.30	0.31
	β^{AIR_i}	0.17	0.20	0.34	0.63	0.71
	β^{CARD_i}	-0.18	-0.15	0.00	0.15	0.19
	$\beta^{TOTAL.EARN_i}$	-0.35	-0.28	-0.07	0.10	0.14
	β^{TENURE_i}	-0.02	-0.02	0.01	0.05	0.06
	β^{ELITE_i}	-0.30	-0.24	-0.07	0.07	0.10
Population means of	$\gamma_i^{BONUS_{ij}}$	-0.58	-0.49	-0.18	-0.02	0.00
$\boldsymbol{\gamma}_i$	$\gamma_i^{WALLET_{ij}}$	-1.30	-1.16	-0.39	0.01	0.07
$(\mu_5 - \mu_{10})$	$\gamma_i^{EARN_{ij}}$	-1.51	-1.40	-0.91	-0.54	-0.49
	$\gamma_i^{SPEND_{ij}}$	-0.33	-0.27	0.05	0.37	0.44
	$\gamma_i^{DEPLETE_{ij}}$	0.02	0.05	0.17	0.30	0.33
	$\gamma_i^{BALANCE_{ij}}$	-0.60	-0.54	-0.23	0.01	0.05

Appendix C: k-medoids probabilistic segmentation

We first apply the k-medoids algorithm to the standardized posterior medians of the consumers' behavioral parameters and plot the consumer segments in Figure 2. This *deterministic* segmentation assigns each consumer to a segment but does not account for the uncertainty in estimating the model parameters. To propagate the posterior uncertainty to the segmentation, we first apply the k-medoids algorithm to each of the 4,000 posterior samples that we obtained from the Bayesian estimation. Next, we aggregate the ensemble of 4,000 different segmentation results to obtain a *probabilistic* segmentation for each consumer. These results are plotted in Figure 9 of the main paper.

The probabilistic segmentation provides additional insights compared to the deterministic segmentation. For instance, the probabilities of consumers belonging to Segment 4 range from 28% to 41% with a median of 34%, reflecting lower confidence in segment assignment compared to Segments 1-3, where consumers have membership probabilities ranging from 27% to 95% with a median of 49%. Consumers in Segment 4 also have median probabilities of memberships in Segment 1 and 2 of 24% and 26%, respectively. This suggests that many consumers have fuzzy memberships among Segments 1, 2, and 4. Firms may then have to consider the behavioral drivers of all three segments if they wish to target Segment 4 consumers.

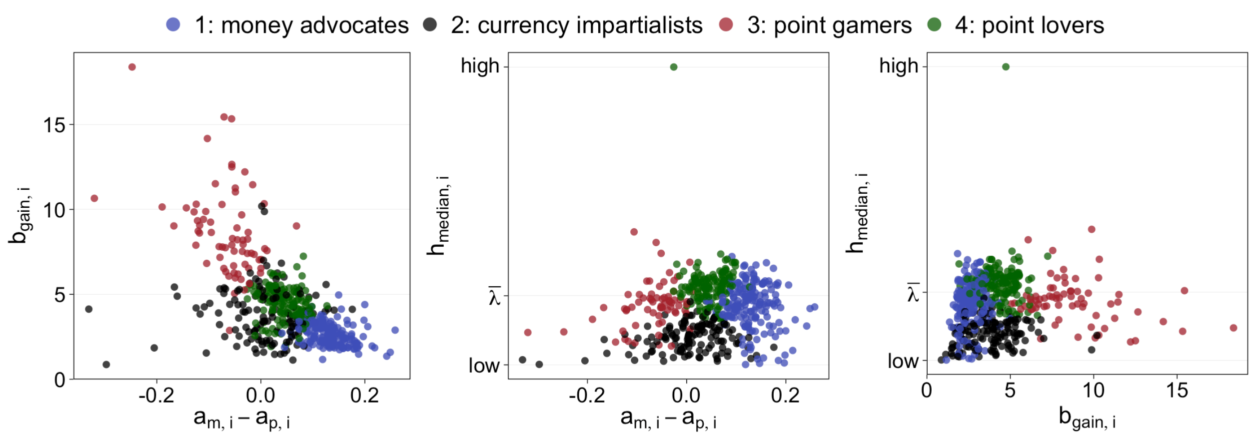


Figure 2 **k-medoids deterministic segmentation of consumers along dimensions of $a_{m,i} - a_{p,i}, b_{gain,i}$ and $h_{median,i}$. Points show consumers' median posterior parameter values.**

This probabilistic segmentation method is similar to bootstrap aggregation clustering procedures, in which a clustering algorithm is applied to different bootstrapped samples before being aggregated. The advantage of the Bayesian framework is that we obtain multiple posterior samples as a natural product without having to perform bootstrapping. Dudoit and Fridlyand (2003) proposes a simple majority vote rule that assigns each observation to the cluster with the highest proportion of assignments across samples. This allows us to cluster consumers to a segment and assign a probability of cluster membership. We also need to undo the cluster label switching across samples, which arises because different permutations of cluster labels or names can result in the same solution to the k-medoids optimization problem. For example, a segment of consumers may be labelled as cluster '1' in one sample but as cluster '2' in another sample, since the labels '1' and '2' are interchangeable to the algorithm and lack any intrinsic meaning. Simply aggregating the cluster labels without undoing the label switching will result in these consumers being assigned to clusters '1' and '2' with 50% probabilities each. When we undo the label switching, we may, for example, relabel cluster '2' to '1' in the second sample such that consumers from this segment consistently belong to cluster '1' both samples. Dudoit and Fridlyand (2003) proposes permuting the labels for each sample such that there is maximum overlap with the cluster assignments of a reference sample, which in their case is the original non-bootstrapped dataset. This is similar to the Equivalence Classes Representative algorithm (Papastamoulis and Iliopoulos 2010, Papastamoulis 2016), originally formulated to address the label switching problem that arises in Bayesian mixture models. For undoing the label switching in our context, we use the deterministic segmentation results from Figure 2 as the reference segmentation.

To conclude, Figure 3 shows the k-medoids deterministic segmentation results for $k = 3, 4, 5$. Setting $k = 4$ roughly splits Segment 3 (from $k = 3$) into two segments, and setting $k = 5$ roughly splits Segment 1 (from $k = 4$) into another two segments. We choose to set $k = 4$, which produces the most distinctive segments and offers meaningful managerial insights.

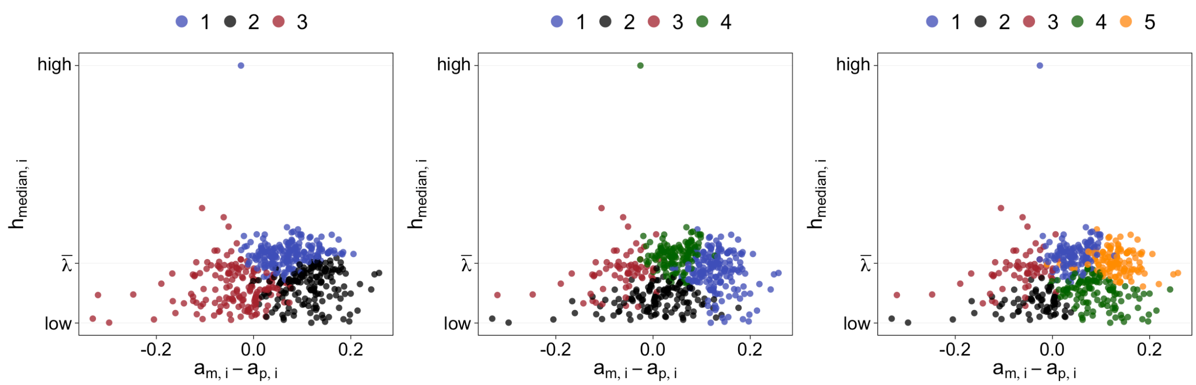


Figure 3 Varying number of segments for k-medoids segmentation.

We can also use the silhouette score criterion to assess the quality of the clustering (Rousseeuw 1987). This is a common method that has been used to assess the clustering quality for brands in marketing (Yang et al. 2022), research articles in operations (Roels and Staats 2021) and firms in finance (Evgeniou et al. 2022). For each consumer i, the silhouette score is defined as $s(i) = \frac{b(i)-a(i)}{\max(b(i),a(i))}$, where $a(i)$ is the mean distance between i and other consumers in the assigned cluster a, and $b(i)$ is the mean distance between i and other consumers in the nearest non-assigned cluster b. The quality of clustering is better when consumer i is similar to other consumers in the assigned cluster, such that $a(i)$ is low, and when consumer i is dissimilar to other consumers in the nearest non-assigned cluster, such that $b(i)$ is high. High $s(i)$ scores therefore indicate that consumer i is well assigned to a cluster. We then compute the mean $s(i)$ across all consumers for the k-medoids algorithm for different numbers of clusters k. The results are shown in Figure 4.

The top two highest clustering scores occur for the 2 and 4 cluster segmentations. The 2 cluster segmentation, however, essentially disregards the segmentation dimension of $h_{median,i}$ and segments consumers along the $a_{m,i} - a_{p,i}$ and $b_{gain,i}$ dimensions only (see Figure 5). On the other hand, the 4 cluster segmentation

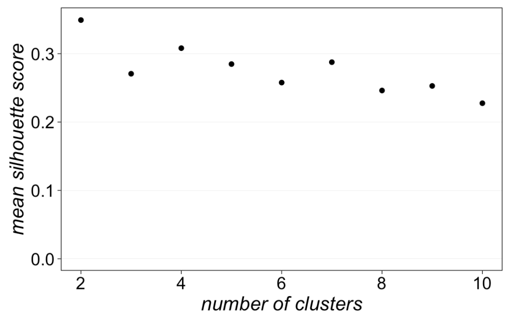


Figure 4 **Silhouette scores of k-medoids segmentation for different number of clusters. Higher scores are better.**

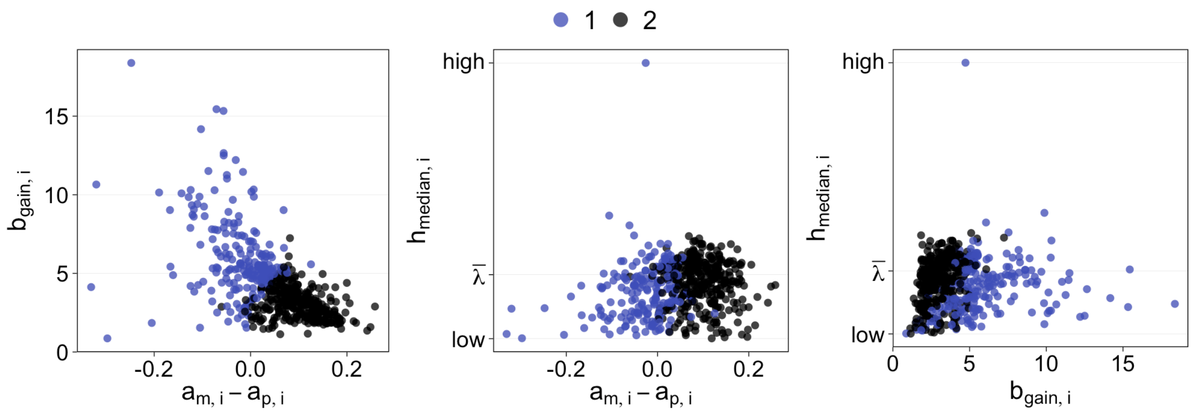


Figure 5 **k-medoids deterministic segmentation with 2 clusters.**

segments consumers along all dimensions, providing more comprehensive managerial insights. Thus, we focus on the 4 cluster result.

Figure 6 shows some observable consumer characteristics for each segment. Currency impartialists and point gamers (segments 2 and 3) have higher and more heterogeneous redemption rates, while money advocates and point lovers (segments 1 and 4) have lower and less varied redemption rates. We discuss this observation in greater detail in Section 6.3 of the main paper. Compared to other segments, a currency impartialist (segment 2) purchases air tickets more frequently, and is more likely to be male and to have elite status. This suggests that the choices of elite status consumers are not driven by good point deals. Rather, they choose between money and points payment by weighing prices more objectively. Point gamers and point lovers (segments 3 and 4) are more likely than other segments to be female. On the other hand, all segments of consumers are very similar in the number of years that they are enrolled in the program, and in their booking windows for ticket redemptions.

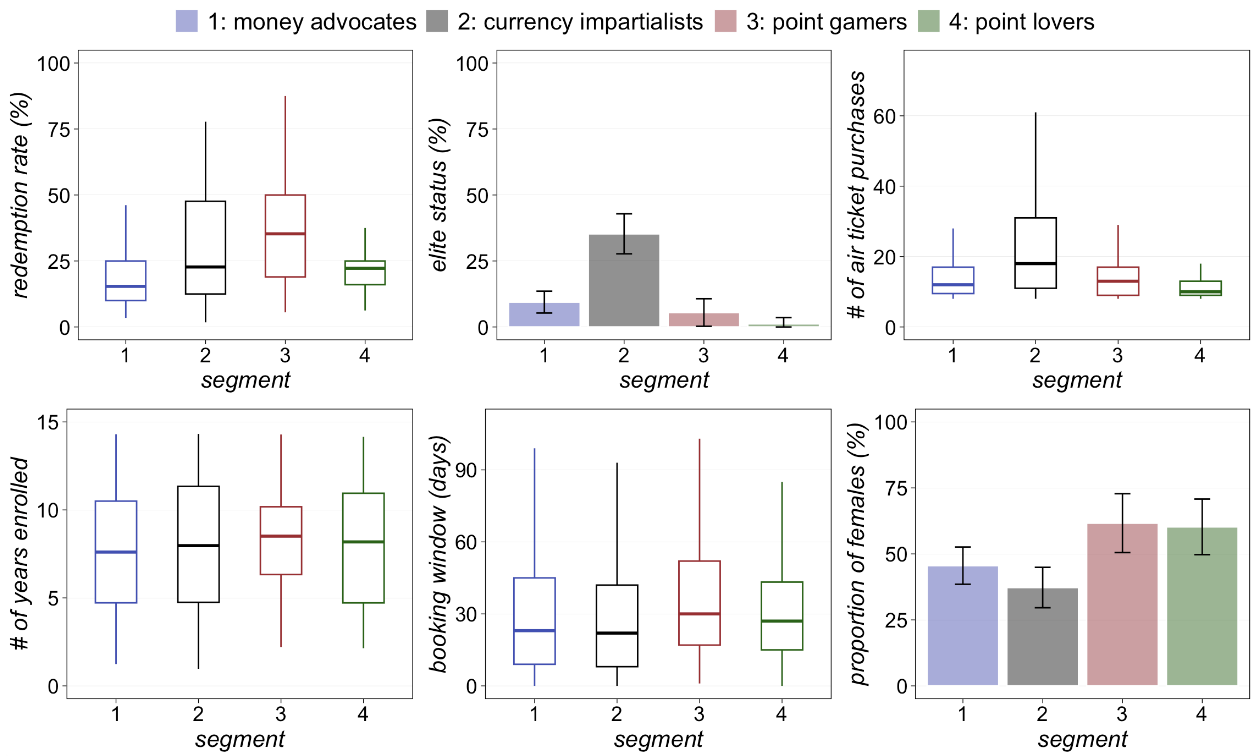


Figure 6 Consumer characteristics for each segment.

Appendix D: Optimal discount levels for different levels of $\eta_m^{indirect}$ and $\eta_p^{indirect}$

In this section, we recompute Figure 9 from the manuscript (which shows the revenue increase of discount policies) for different levels of the indirect revenue impacts of money $\eta_m^{indirect}$ and points $\eta_p^{indirect}$. The specific levels of optimal discounts and revenue increases vary with the firm's assessed levels of $\eta_m^{indirect}$ and $\eta_p^{indirect}$ but the qualitative conclusions remain robust and unchanged.

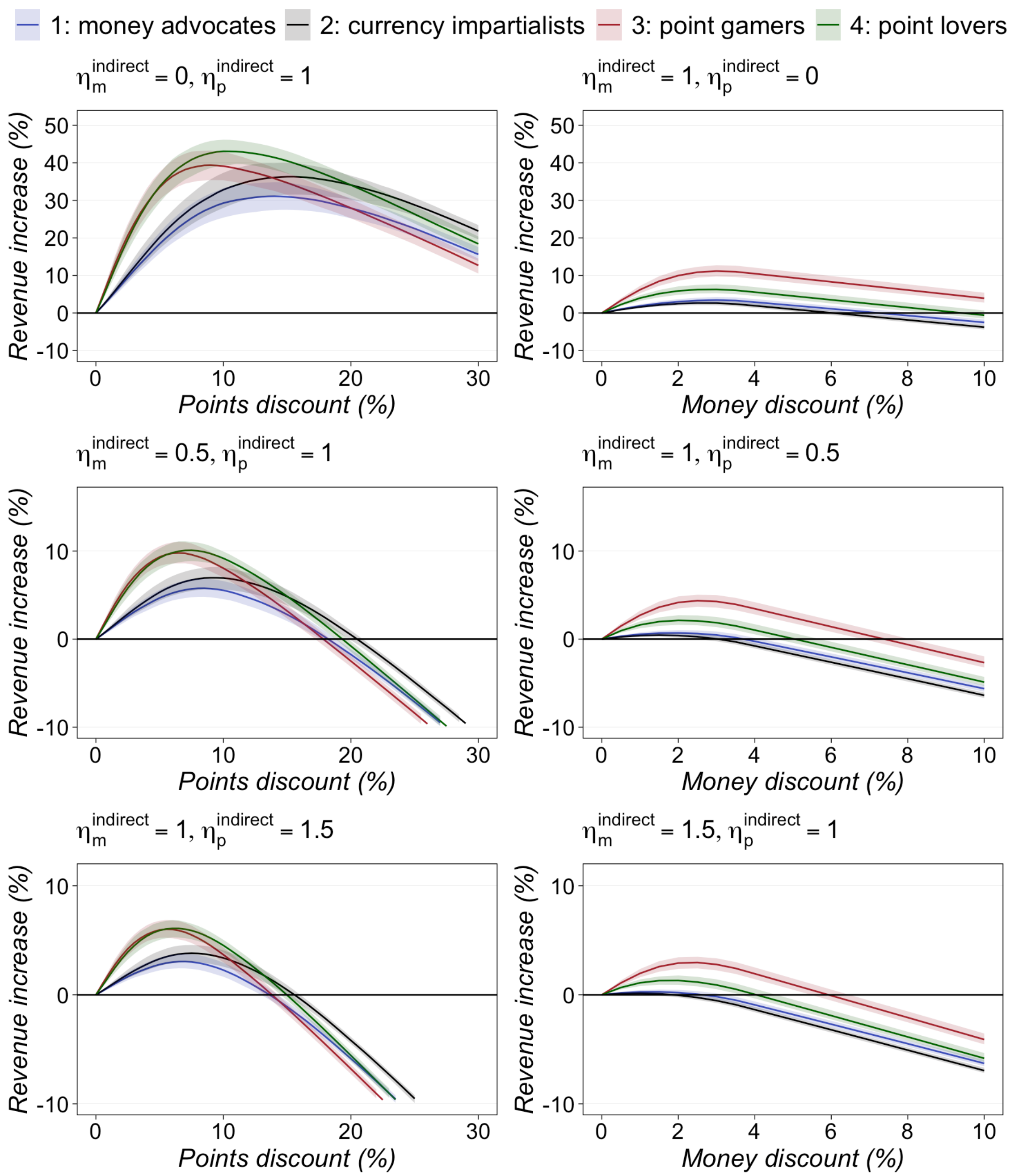


Figure 7 Percentage revenue increase from the implementation of points and money discount policies, compared to the baseline of no discounts.

Appendix E: Robustness checks

E.1. Proxy of booking dates

We proxied consumers' booking dates for flights paid with money using their known booking and departure dates for flights paid with points. Specifically, we calculate the median time between the booking and departure dates (booking window) of all redemptions made by the consumer. We then obtain their booking dates for tickets purchased with money by subtracting this median booking window from the observed departure dates. However, a consumer may systematically book flights that are paid with money further from or closer to the departure date, compared to flights that are paid with points. This proxy of booking dates may therefore result in systematic errors relative to the true values. Even if a consumer does not exhibit systematic differences in booking windows between flights paid with money and flights paid with points, proxying all of the consumers' booking windows for flights paid with money with a single value will still result in errors relative to the true values.

We try to address this problem next. We construct four adjusted datasets in which the proxied booking dates are either systematically shifted forward or backward, or are varied across transactions. Specifically, the number of days by which the proxied booking dates are changed for these adjusted datasets are shown in Figure 8.

1. **Booking dates shifted earlier**: For each booking date that is proxied with the median booking window of redemptions, we increase the booking window by 50% for all booking dates that are proxied. This describes a scenario in which consumers systematically book tickets paid with money earlier than tickets paid with points. We then reconstruct the dataset by revalidating the decision occasions and recomputing the decision level explanatory variables in $\boldsymbol{v}_{ij}$. A decision occasion is considered valid if consumers' point balances at the date of booking is greater than or equal to the point prices, such that consumers are making payment choices between money and points.

2. **Booking dates shifted later**: Same as (1), except that we decrease the booking window by 50% for all booking dates that are proxied. This describes a scenario in which consumers systematically book tickets paid with money later than tickets paid with points.

3. **Booking dates + error**: Same as (1), except that we either increase or decrease the booking window by 50% (with the direction randomly generated). This shifts each proxied booking date backward or forward, respectively.

4. **Predicted booking dates**: We train an elastic net prediction model for booking windows using the set of tickets paid with points. We include the prediction variables: flight route, flight date at the year-week level, money price, point price, cabin class, membership tenure, age, gender, elite status and state of residence. We use the R package *glmnet*, which performs 10-fold cross validation by default and returns an optimal regularization or tuning parameter. We run this procedure for different levels of the elastic net parameter, $\alpha \in \{0, 0.25, 0.5, 0.75, 1\}$. This controls the elastic net mix between lasso regression ($\alpha = 1$) and ridge regression ($\alpha = 0$). We find that $\alpha = 0.5$ produces the lowest mean absolute error during cross validation. Therefore, we predict the booking windows (and compute new booking

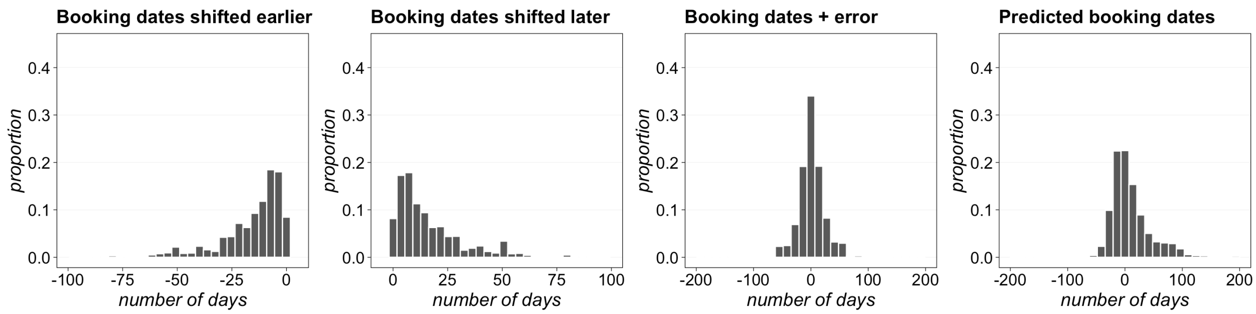


Figure 8 Number of days by which the proxied booking dates are shifted.

dates) for tickets paid with money using the elastic net model with $\alpha = 0.5$. We then reconstruct the dataset in the same way as (1).

We estimate the full model for each of the 4 adjusted datasets in which the booking dates are changed, and compare the results with the main results and findings from the manuscript. Figures 9-11 show how the results would change under scenarios (1)-(4) and confirm that the results remain robust under different types of perturbations of the proxied booking dates.

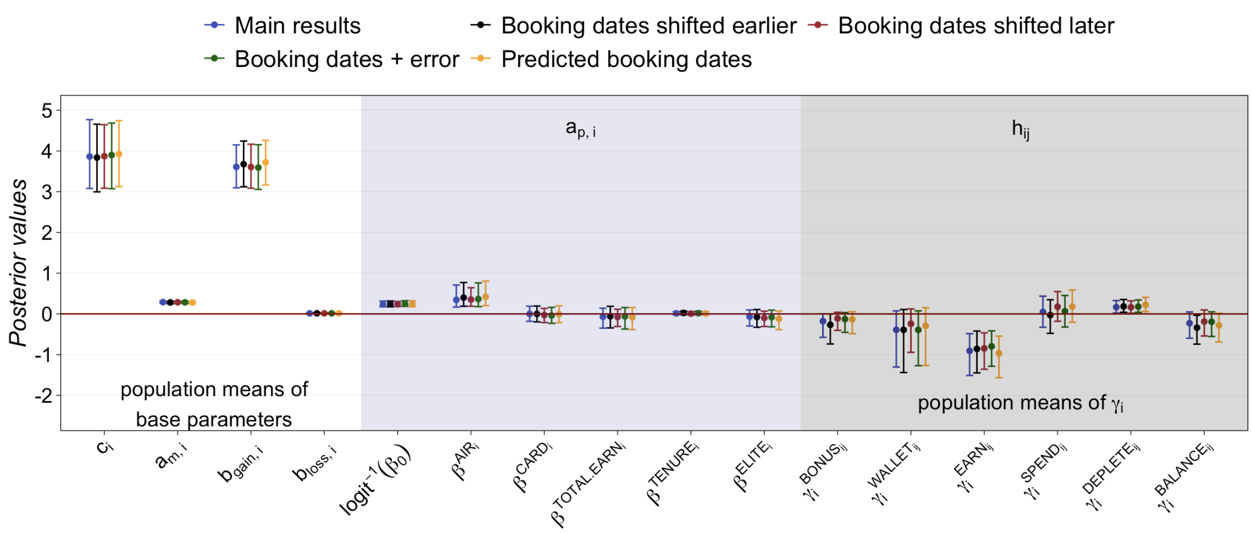


Figure 9 The 95% credible intervals of the population means in the main results, compared with those from the adjusted datasets.

To understand these results further, observe that changes in the proxied booking dates can alter the model results through two avenues. First, it may change the validity of a decision occasion. In particular, the proxy of booking dates for tickets that are paid with money may change the ordering of a consumer's transactions. A consumer may accrue points from various sources over time, face a decision to pay with money or points for a flight ticket, choose to pay with money, and accrue the rewarded points from the money payment in future when they complete the flight. We only observe consumers' point balances and not their money balances, therefore we cannot determine with certainty whether their money payment decisions have been inserted in the right order among their point transactions. If money payment decisions are inserted in the

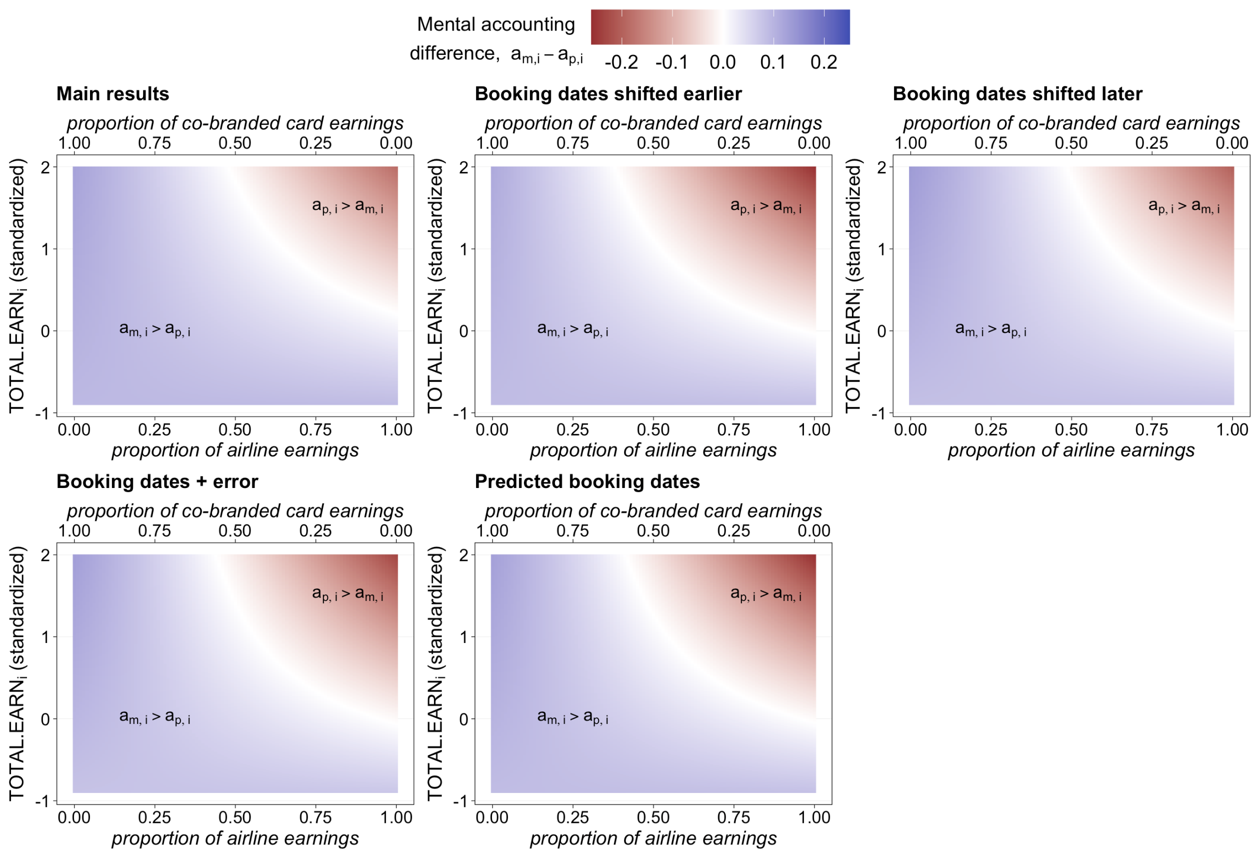


Figure 10 Effects of point earning source variations on the mean of the mental accounting differences.

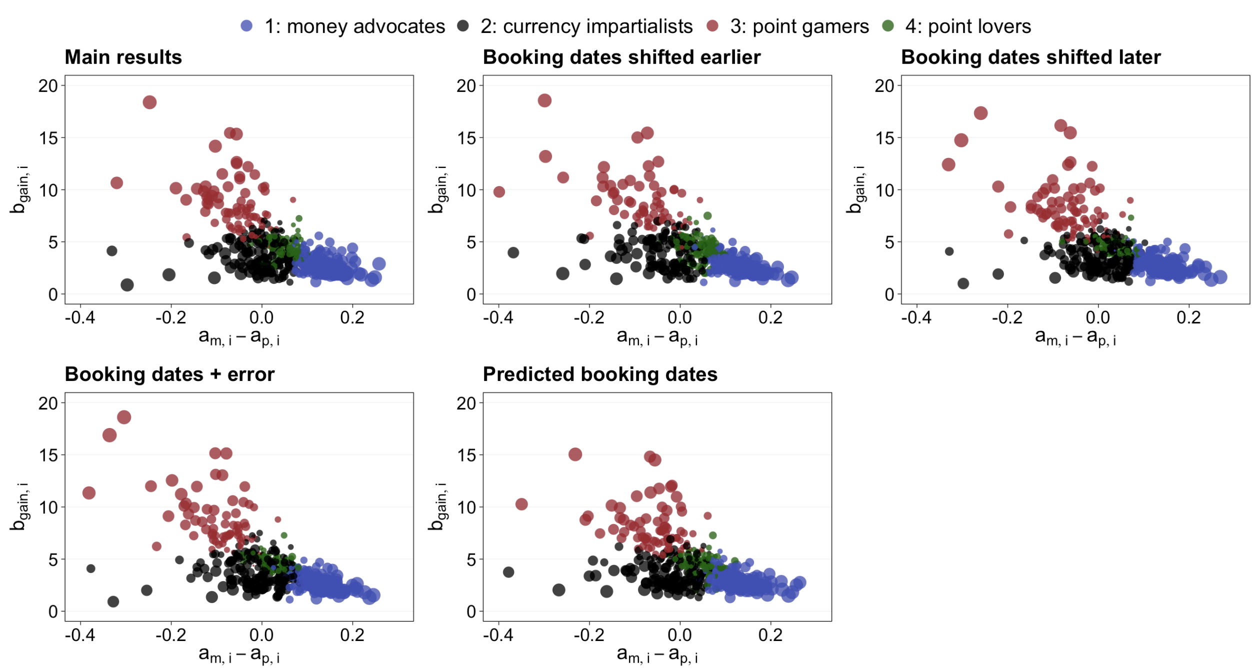


Figure 11 k-medoids probabilistic segmentation of consumers.

wrong order, it can lead to incorrect point balances, affecting the determination of a valid decision occasion. Recall that a decision occasion is valid if consumers' point balances at the date of booking is greater than or equal to the point prices, such that consumers are making payment choices between money and points.

This error, however, is mitigated by the fact that there is a sizeable buffer window within which the proxied booking date can be shifted without invalidating the decision occasion, i.e., a consumer's point balance within this buffer window is always greater than or equal to the point price. As shown in Figure 12, for 39% of decision occasions in which a money payment is chosen and the booking date is proxied, the decision occasion will never be invalidated because the consumer has sufficiently large point balances throughout the data period. For the remaining 61% of decision occasions, the booking date can be shifted by an average of 138 days, or 4.6 months, before the decision occasion is rendered invalid. This represents a very large shift. For comparison, the average proxied booking window for tickets paid with money in the main dataset is only 34 days.[2]

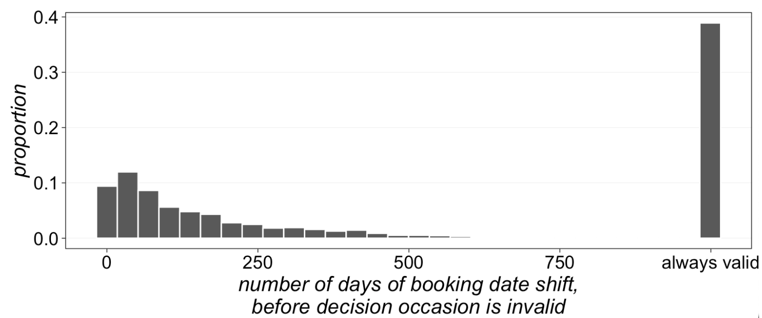


Figure 12 Number of days by which the proxied booking date for a ticket purchased with money can be shifted, before the decision occasion is rendered invalid.

Second, if the booking dates of tickets paid with money are shifted, the explanatory variables in $\boldsymbol{v}_{ij}$, which are computed up to the point of decision occasion, may also change. Fortunately, these variables are not significantly affected by changes made to the proxied booking dates in the three adjusted datasets, as shown in Tables 2 and 3.

Table 2 Mean absolute differences between explanatory variables in $\boldsymbol{v}_{ij}$ in the adjusted and original datasets.

	$BONUS_{ij}$	$WALLET_{ij}$	$EARN_{ij}$	$SPEND_{ij}$	$DEPLETE_{ij}$	$BALANCE_{ij}$
Booking dates shifted earlier	0.03	0.02	0.09	0.04	0.07	0.05
Booking dates shifted later	0.03	0.02	0.09	0.06	0.08	0.05
Booking dates + error	0.03	0.01	0.08	0.05	0.08	0.05
Predicted booking dates	0.04	0.02	0.10	0.07	0.10	0.06

[2] We also perform a robustness check in which we only include "likely valid" decision occasions in our dataset. These are defined as those occasions for which a booking date shift of greater than 6 months or 180 days is required to invalidate the decision. This procedure will skew the dataset towards those decision occasions with high point balances, since these are more likely to be valid. For the dataset used in the main results, the mean point balance is about 60,000. For the dataset used in this robustness check, the mean point balance is about 78,000. In spite of this, we find that our main results remain robust.

Table 3 Correlation between explanatory variables in $\boldsymbol{v}_{ij}$ in the adjusted and original datasets.

	$BONUS_{ij}$	$WALLET_{ij}$	$EARN_{ij}$	$SPEND_{ij}$	$DEPLETE_{ij}$	$BALANCE_{ij}$
Booking dates shifted earlier	0.96	0.97	0.99	0.97	0.94	0.99
Booking dates shifted later	0.96	0.98	0.95	0.97	0.94	0.99
Booking dates + error	0.96	0.97	0.99	0.98	0.94	0.99
Predicted booking dates	0.96	0.98	0.99	0.96	0.91	0.99

To make this specific, the mean absolute differences between the recomputed explanatory variables in $\boldsymbol{v}_{ij}$ of the adjusted datasets and those of the original dataset are shown in Table 2. The variables $BONUS_{ij}$, $WALLET_{ij}$, $EARN_{ij}$, $SPEND_{ij}$ and $BALANCE_{ij}$ are standardized such that they have mean 0 and standard deviation of 1. The mean absolute changes in these variables as a result of changing the proxied booking dates are between 0.01 and 0.10, which are small relative to the scale of the variables. $DEPLETE_{ij}$ is not standardized but is constructed such that its range is between -2 and 1. Its mean absolute change is no more than 0.10, which is again small relative to the scale of the variable. In addition, Table 3 shows the correlations between the recomputed explanatory variables in $\boldsymbol{v}_{ij}$ of the adjusted datasets and those of the original dataset. All correlations are above 0.91. If, for example, the mean monthly bonus point earnings is high in the original dataset, it is still the case even after the booking dates are changed.

In summary, the main sources of errors that can affect the results from the proxy of booking dates are mitigated in various ways. Changing the proxied booking dates only result in small changes in the model inputs. Therefore, the results remain robust to errors that may arise from the proxy of booking dates for tickets purchased with money.

E.2. Proxy of point prices (exchange rates)

In the main results, we match each money payment decision occasion with point payment decision occasions using the matching dimensions of origin-destination route and departure dates with a ±2 week window. We then compute the average points exchange rate $\hat{\lambda}$ of the matched redemptions and set the points price of the money payment decision occasion to be $p_{ij} = m_{ij}/\hat{\lambda}$. We choose to match on these 2 dimensions of route and departure date because this leads to relatively good performance and high success rate in matching (99.5%). The latter point is particularly important because it ensures that observations do not drop out of the final dataset due to a failure to proxy their exchange rates and points price. Observations that drop out will tend to be less popular flights (as defined by the matching dimensions) and their omission can result in a dataset skewed towards a particular demographic of flights.

In this section, we will compare the original matching method with an alternative that additionally matches on the dimensions of booking date (±2 week window) and cabin class (business class vs economy class). As the percentage of successfully matched observations with this alternative is a lot lower (75.1%), we also consider the alternative method with a larger matching window of ±1 month for booking dates.

We compare these matching methods on the redemptions data where both money and point prices are known. In particular, we test how well the matching algorithm performs in estimating the exchange rate and point price for each observation in this test set, using only the remaining $n-1$ observations in the test set, a principle that is analogous to leave-one-out cross validation. When we assess the performance of the alternative matching method, note that even though we know the true booking dates in this test set, we also proxy the booking dates for each observation using the remaining $n-1$ observations in the same manner that we do for tickets purchased with money in the main dataset.

Table 4 Predictive statistics of matching algorithms using test set of point redemptions. Mean absolute percentage error (MAPE) and mean percentage error (MPE) values are calculated based on the common set of matchable data points of the alternative matching algorithm. "✓" indicates that matching is performed on the specified dimension.

	Route	Flight date	Cabin class	Booking date	% matched	MAPE (rate)	MAPE (points)	MPE (rate)	MPE (points)
Original matching	✓	✓			99.5	0.039	0.036	0.008	-0.003
Alternative matching	✓	✓	✓	✓	75.1	0.038	0.036	0.009	-0.003
Alternative matching (large window)	✓	✓	✓	✓	88.2	0.037	0.035	0.008	-0.003

Table 5 Summary statistics of datasets obtained with different matching methods.

	# of consumers	# of decision occasions	% redemptions	Mean purchases per consumer
Original matching	500	8,839	21.5	17.7
Alternative matching	391	6,447	25.3	16.5
Alternative matching (large window)	391	7,157	22.7	18.3

Table 4 shows the out-of-sample predictive accuracies of the competing methods. In addition, Table 5 offers summary statistics of the datasets after the proxying has been carried out. Together, these two tables provide us with several observations. First, the original matching performs well, with mean absolute percentage errors of less than 4%. It performs very close to the alternative matching methods, which match on more dimensions. However, the alternative matching methods have a significant flaw in that the percentage of successful matches and proxies decreases significantly to as low as 75%. This results in a skewed dataset with inflated redemption rates (as seen in Table 5) because we have to drop observations of tickets purchased with money for which we cannot proxy the exchange rates and point prices.

Figure 13 compares the posterior population mean estimates between the original and alternative matching methods. The posterior population mean estimates remain stable with the exception of $b_{gain,i}$, for which median posterior estimates (indicated by the solid dots) vary between 1.6 and 3.6. To interpret this parameter, we express it in terms of the effect on redemption propensities. These median posterior estimates translate to an average increase in the probability of points payment of between 2.4% and 3.0%[3] for every 0.01 cent

[3] This effect size of $b_{gain,i}$ is computed at the median point price with the points exchange rate set to the reference level of $\bar{\lambda}$. $\log(h_{ij})$ and $\text{logit}(a_{p,i})$ are set to the posterior medians of γ_0 and β_0, respectively.

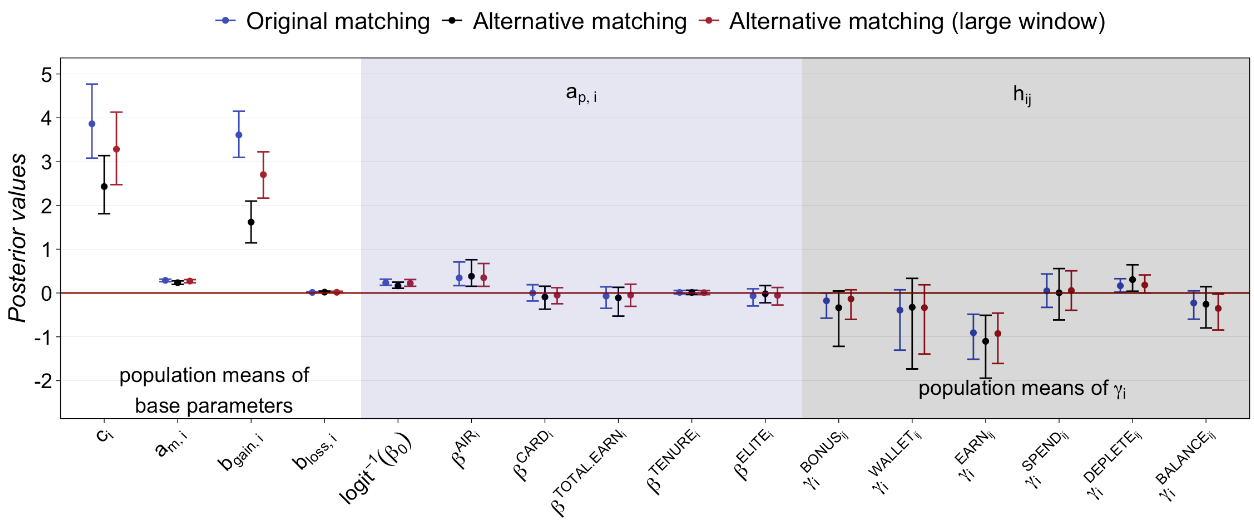


Figure 13 The 95% credible intervals of the population means, for the original and alternative matching methods used in the proxy of point prices (exchange rates).

per point increase in exchange rates[4], which represent a tight range of effect sizes. The posterior estimates of $b_{gain,i}$ also remain significantly larger relative to those of $b_{loss,i}$. Therefore, the qualitative conclusions of the results remain robust and unchanged regardless of the matching method used.

Figure 14 shows that the qualitative conclusions on the joint effects of point earning level and point earning sources remain robust. When the alternative matching method is used, there is a visibly larger region in which the mental accounting parameter for points is lower than that for money for a hypothetical consumer across a grid of point earning profiles. This is consistent with the skewed dataset that is produced with the alternative matching method (recall Table 5). Given that this matching method drops many money purchases, it leads to a dataset with a higher point redemption rate. Consumers are therefore estimated to value points less resulting in them redeeming more.

Figure 15 shows that the qualitative conclusions regarding consumer heterogeneity and segmentation also remain robust and unchanged. When the alternative matching method is used, a larger proportion of consumers have lower mental accounting parameter for points relative to money (i.e., they have positive $a_{m,i} - a_{p,i}$ values). This is again consistent with the skewed dataset having a higher point redemption rate.

Next, we check the robustness of the proxy method by adding error and noise to the proxied exchange rates in the following ways:

1. **Rates + error:** In the first part of the analysis in this section, we test the performance of the matching algorithm on the test set of airline ticket point redemptions. The histogram of prediction errors are shown in Panel A of Figure 16. Under this "rates + error" scheme, we assume that this error distribution will remain the same when we proxy money payment exchange rates with redemption data points. Specifically, for tickets purchased with money, we shift the proxied rates by percentages in accordance

[4] As reference, most airline loyalty points are valued around 1 - 2 cents per point. In the dataset, where most point redemptions occur above a reference exchange rate level $\bar{\lambda}$, exchange rates are within a range of $[\bar{\lambda} - 0.41, \bar{\lambda} + 0.05]$ cents per point.

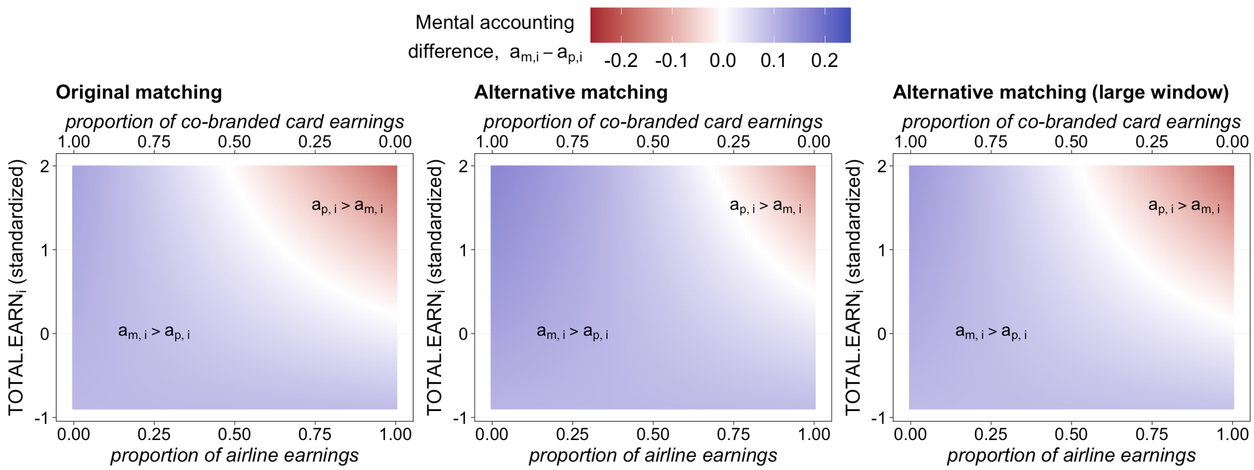


Figure 14 **Effects of point earning source variations on the mean of the mental accounting differences.**

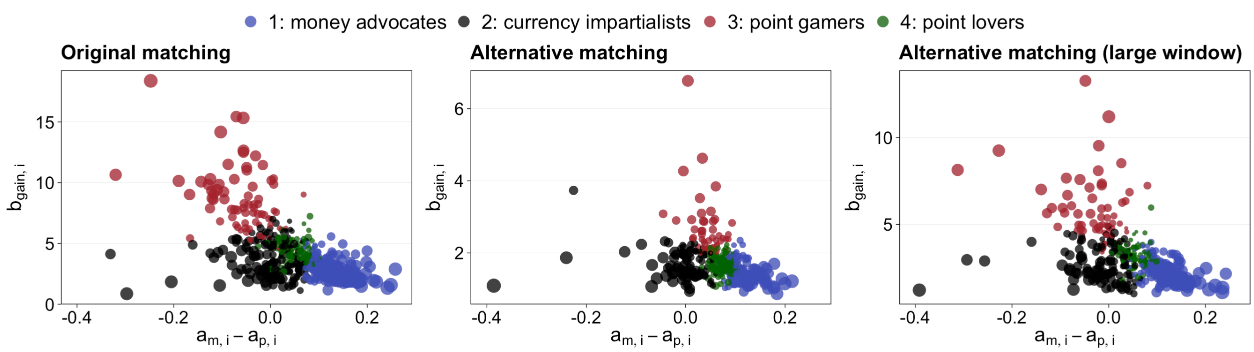


Figure 15 **k-medoids probabilistic segmentation of consumers. The scales for $b_{gain,i}$ are kept different in order to display the segmentation.**

with this distribution of prediction errors when we proxied redemption exchange rates with redemption data points. We also restrict the shifted exchange rates to be within the minimum and maximum observed exchange rates in the dataset. The resultant percentage shifts that are applied are shown in Panel B of Figure 16. The point prices are then computed based on the shifted exchange rates, the validities of the decision occasions are reassessed, and an adjusted dataset is constructed.

2. **Rates + noise:** We change the proxied exchange rates by a percentage that is simulated as $N(0, 0.04^2)$. The magnitude of the standard error of the percentage change is set equal to the mean absolute percentage error of the predictive performance for the original matching method, as tested on the test set in the first part of the analysis in this section (recall Table 4). The resultant percentage shifts that are applied are shown in Panel C of Figure 16. In the "rates + error" scheme, there are many percentage shifts that are close to zero, and there are some large negative shifts. In the "rates + noise" scheme, the percentage shifts are more symmetrically distributed.

Figure 17 compares the posterior population mean estimates between the original and adjusted datasets. The posterior estimates remain robust to added error or noise. In particular, the posterior median estimates for the population mean of $b_{gain,i}$ vary between 1.9 and 3.6, translating to average increases in the probability

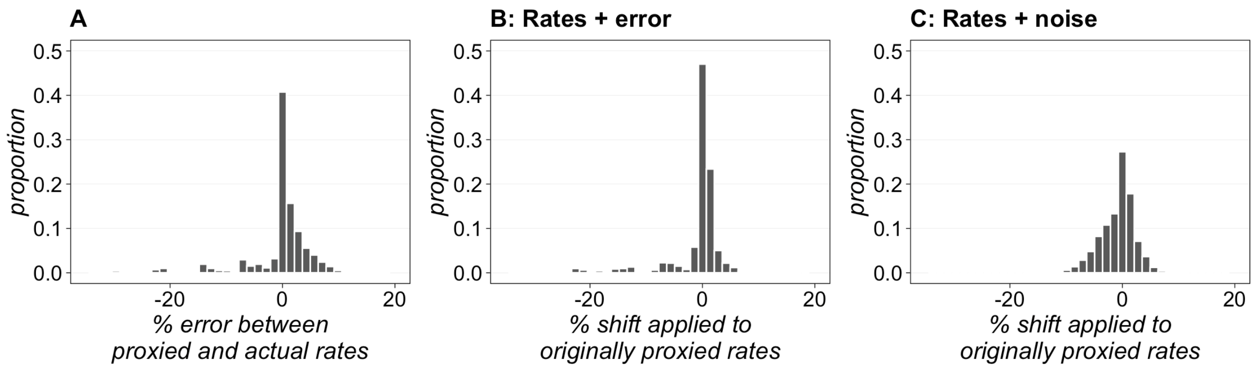


Figure 16 **Panel A. Prediction errors of original matching, tested on test set of airline ticket point redemptions. Panel B. Errors applied to originally proxied exchange rates. Panel C. Noise applied to originally proxied exchange rates.**

of points payment of between 2.6% and 3.3%[5] for every 0.01 cent per point increase in exchange rates. This represents a tight range of effect sizes.

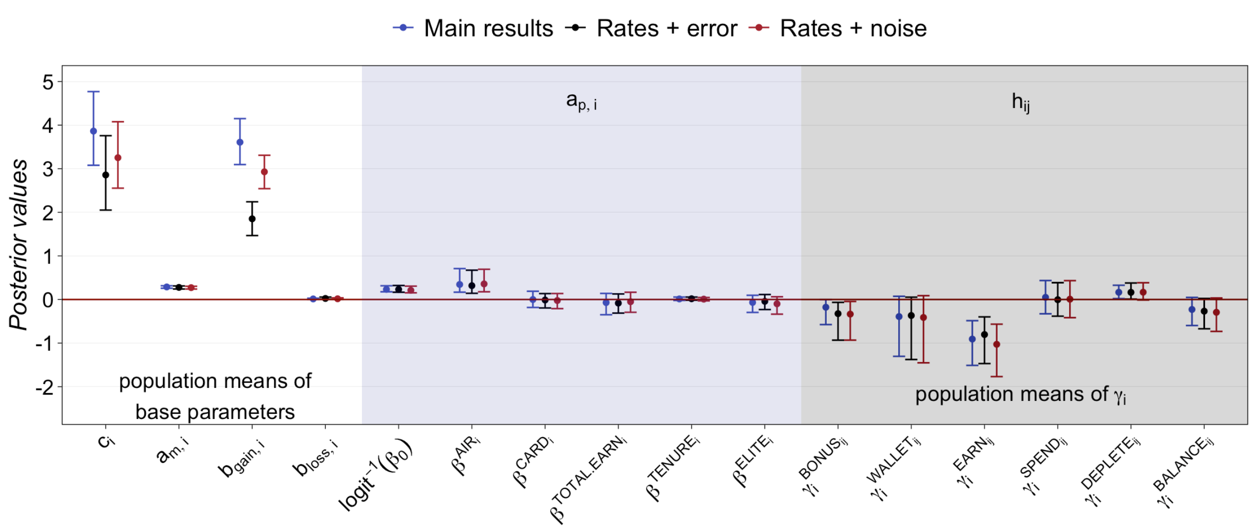


Figure 17 **The 95% credible intervals of the population means in the main results, compared with those from the adjusted datasets in which error and noise are added to the proxied exchange rates.**

The joint effects of point earning level and source on mental accounting differences in Figure 18 and the segmentation of consumers in Figure 19 also remain robust to added error or noise, and the qualitative conclusions remain unchanged.

In summary, we have adopted a parsimonious method to recover unobserved point prices (exchange rates) for money payment decision occasions in the main results. This method performs well when applied on a

[5] This effect size of $b_{gain,i}$ is computed at the median point price with the points exchange rate set to the reference level of $\bar{\lambda}$. $\log(h_{ij})$ and $\text{logit}(a_{p,i})$ are set to the posterior medians of γ_0 and β_0, respectively. Note that the effect size of 3.3% corresponds to the “Rates + noise” scheme in which the posterior median estimate of the population mean of $b_{gain,i}$ is 2.9.

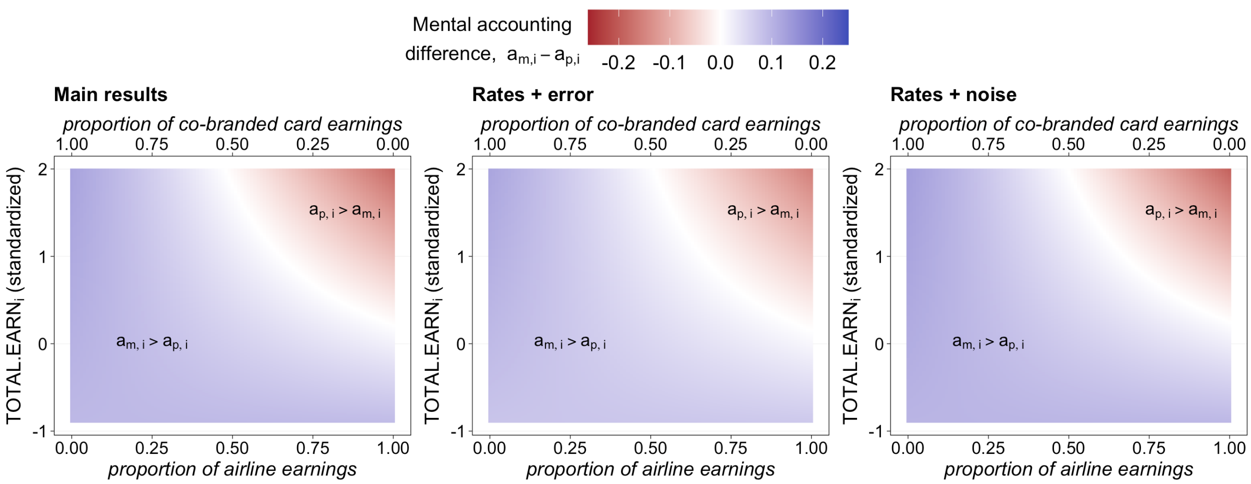


Figure 18 **Effects of point earning source variations on the mean of the mental accounting differences.**

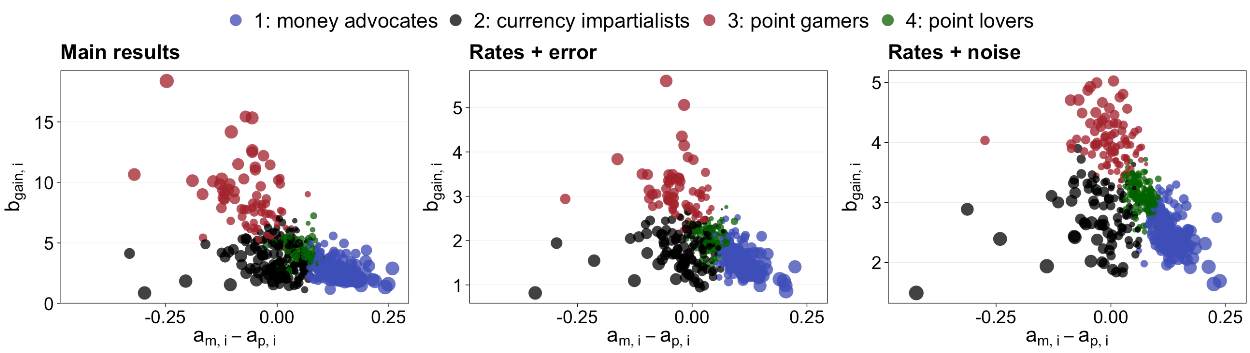


Figure 19 **k-medoids probabilistic segmentation of consumers. The scales for $b_{gain,i}$ are kept different in order to display the segmentation.**

test set of redemption decisions with known prices, and also performs very close to other matching methods that attempt to match observations on more dimensions. Even when we artificially add errors and noise to the proxied variables, our results remain robust.

E.3. Consideration of consumers' flight choices with a nested logit model

If the point exchange rate for one flight is high (resulting in a relatively attractive point price) and the point exchange rate for another flight is low (resulting in a relatively attractive money price), a consumer may compare the point price of the first flight to the money price of the second flight. This will contravene our model in which consumers compare the money and point prices within flights. This concern is mitigated by model free evidence that we show in Figure 20. In this figure, we construct a coarse consideration set or flight menu for each flight purchase observation, with a flight menu defined as a collection of flights on the same route that departs within ±1 day of one another. For each flight menu, we compute the maximum difference in exchange rates among the flights, and plot the histogram in Figure 20. We observe that for the majority of flight menus, the maximum difference in exchange rates between different flights is very small. Consumers will then be choosing between flights with similar point exchange rates, giving them little reason to compare prices across flights.

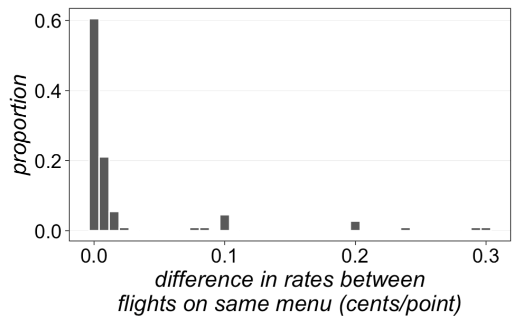


Figure 20 Maximum difference in exchange rates between different flights on the same flight menu. For reference, most airline loyalty points are valued around 1 - 2 cents per point.

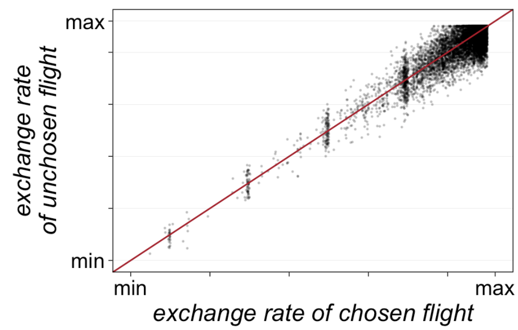


Figure 21 Consumers may be choosing between 2 flights with different point exchange rates.

Despite this model free evidence, we still explicitly assess a nested logit choice model in which consumers jointly choose a flight and currency payment option. We show that the main results remain robust. With the dataset, we cannot satisfactorily reconstruct the consideration set of flights for all observations because a consumer may be considering a flight that is not chosen by any other consumer and hence is not in the dataset. Therefore, we simulate flight level data to demonstrate how incorporating flight choice may affect the results. For simplicity and demonstration purposes, we model consumers as choosing only between two flights during each decision occasion $j \in \{1, ..., J\}$. Suppose that during decision occasion j consumer i considers two flights, A_{ij} and B_{ij}, and chooses flight $f_{ij} \in \{A_{ij}, B_{ij}\}$.

We currently observe data that describes the price variables for a chosen flight. We therefore need to augment this observed data with simulated data that describes flight level data as well as the price variables for the unchosen but considered flight. We simulate one flight level variable for each flight, namely $ATTRACT_{f_{ij}} \sim U(0,1)$. This variable indicates the overall attractiveness of a flight, with higher values denoting greater attractiveness, and may include factors such as flight duration and departure and arrival times. We then set the chosen flight as the flight with the higher attractiveness variable with 70% probability. This way, the consumer chooses the more attractive flight most of the time but not always. We already observe $m_{ij}^{chosen}, p_{ij}^{chosen}, \lambda_{ij}^{chosen}$ and r_{ij}^{chosen} for the chosen flight, i.e., all entries in the dataset belong to chosen flights. But given that we do not observe these price variables for unchosen flights, we simulate them as follows:

$$
\begin{aligned}
m_{ij}^{unchosen} &= m_{ij}^{chosen}(1+\zeta_m), \\
\lambda_{ij}^{unchosen} &= \lambda_{ij}^{chosen}(1+\zeta_\lambda), \\
\zeta_m &\sim N(0, 0.02^2), \\
\zeta_\lambda &\sim N(0, 0.02^2), \\
p_{ij}^{unchosen} &= \frac{m_{ij}^{unchosen}}{\lambda_{ij}^{unchosen}}, \\
r_{ij}^{unchosen} &= \frac{r_{ij}^{chosen}}{m_{ij}^{chosen}} m_{ij}^{unchosen}.
\end{aligned}
$$

Note that the simulation of prices for the unchosen flight involves the addition of noise to the prices of the chosen flight. Therefore, consumers are simulated to potentially choose between flights that have different exchange rates. These differences are illustrated in Figure 21.

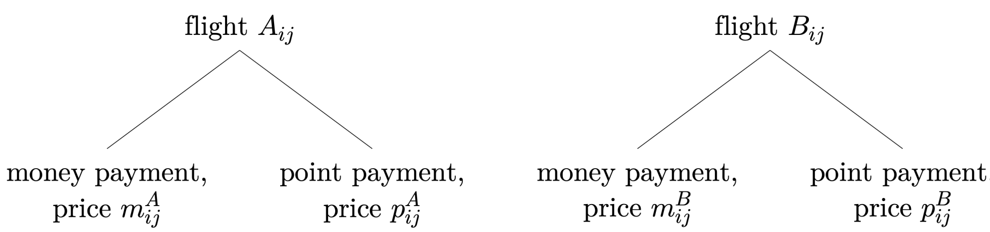


Figure 22 Nested structure of joint flight and currency payment choice.

Figure 22 shows the schematic of the nested logit choice model. Consumers choose from 4 options, formed by all combinations of two currency choices and two flight choices, based on a nested structure that allows currency choices with a common flight to have correlated utilities. The joint probability of consumer i choosing flight f_{ij} and currency payment $c \in \{point, money\}$ during decision occasion j, $\mathbb{P}(f_{ij}, c)$, can be decomposed as $\mathbb{P}(f_{ij}, c) = \mathbb{P}(c|f_{ij})\mathbb{P}(f_{ij})$. The conditional probability of currency choice conditional on flight choice, $\mathbb{P}(c|f_{ij})$, is exactly that specified in the main model in the manuscript[6]. Following standard specifications in the literature (see, e.g., Train 2009, Ben-Akiva and Lerman 1985), we model the marginal probability of flight choice, $\mathbb{P}(f_{ij})$ as follows:

$$
\begin{aligned}
\tilde{y}_{ij} &\sim \text{Bernoulli}(\tilde{q}_{ij}),\\
\text{logit}(\tilde{q}_{ij}) &= \left(U_{ij}^A + IV_{ij}^A - U_{ij}^B - IV_{ij}^B\right)\phi\\
&= \Big(\kappa_0 + \kappa^{ATTRACT} \times ATTRACT_{ij}^A + IV_{ij}^A\\
&\qquad - \kappa^{ATTRACT} \times ATTRACT_{ij}^B - IV_{ij}^B\Big)\phi,\\
IV_{ij}^f &= \log\left(\exp(U_{ij}^{point,f}) + \exp(U_{ij}^{money,f})\right),\\
\kappa_0 &\sim N(0, 5^2),\\
\kappa^{ATTRACT} &\sim N(0, 5^2),\\
g(\phi) &= \begin{cases} 0 & \text{, if } \phi \le 0,\\ 2\phi\exp(-\phi^2) & \text{, if } \phi > 0. \end{cases}
\end{aligned}
$$

Here, $\tilde{y}_{ij} \in \{0, 1\}$ denotes the observed flight choice f_{ij} of consumer i during decision occasion j, with $\tilde{y}_{ij} = 1$ denoting a choice of flight A_{ij} and $\tilde{y}_{ij} = 0$ denoting a choice of flight B_{ij}. $\tilde{q}_{ij}$ denotes the probability of choosing flight A_{ij}. We have specified a simple linear model for the systematic utilities of flight choice U_{ij}^f, comprising an alternative specific constant κ_0 and a single explanatory variable with coefficient $\kappa^{ATTRACT}$.

The inclusive value variable (IV_{ij}^f) is the log sum of exponentials of the utilities of currency payment choices for flight f_{ij}. Intuitively, this conveys a measure of the expected worth of a set of currency payment

[6] The notation we have used in the manuscript is q_{ij} for the probability of point payment.

alternatives for flight f_{ij} (Ben-Akiva and Lerman 1985). This also serves as a link between the upper level flight choice model and the lower level currency choice model, since IV_{ij}^{f} accounts for the utility $U_{ij}^{point,f}$ of paying by points for flight f_{ij}, and the utility $U_{ij}^{money,f}$ of paying by money for the same flight f_{ij}, both of which are specified in the main model in the manuscript[7]. Therefore, if the utility of paying by point or money is high for a particular flight f_{ij}, a consumer is more likely to choose flight f_{ij}. Following Lahiri and Gao (2002), we set a generalized exponential prior for the estimation of the scale parameter ϕ[8], and we complete the setup with relatively diffuse priors for the coefficients κ_0 and $\kappa^{ATTRACT}$.

Next, Figure 23 shows the posterior population level estimates of the nested logit model, after jointly estimating all the parameters of the lower level currency choice model and the upper level flight choice model. The posterior estimates of the two lower level models remain very similar. The reason for this stability is that a nested logit model can also be estimated via sequential estimation: Under such an approach, we first estimate the lower level currency choice model in isolation (which we have done in the main results), use these estimates that we have obtained to construct the inclusive value IV_{ij}^{f}, and finally estimate the upper level flight choice model. Such estimates from sequential estimation have been shown to be consistent (Ben-Akiva and Lerman 1985).

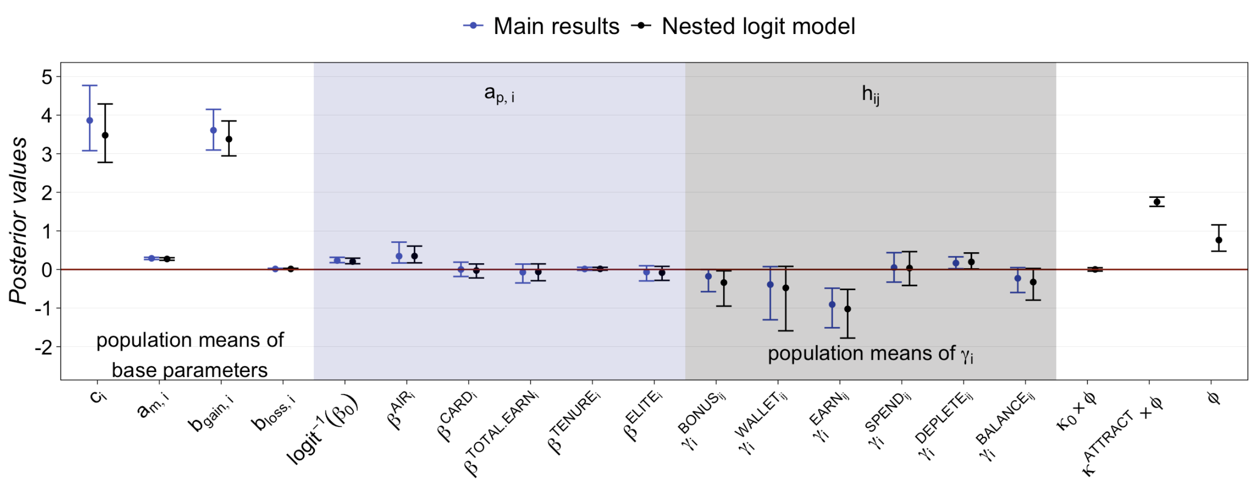


Figure 23 The 95% credible intervals of the population means in the main results, compared with those from the nested logit model.

The alternative specific constant, $\kappa_0 \times \phi$, has a posterior credible interval that is close to and includes zero. This is consistent with the augmented data as we did not simulate any systematic preference for flight

[7] The notation we have used in the manuscript is U_{ij}^{1} and U_{ij}^{0} for the utility of paying by points and money, respectively, conditional on flight choice.

[8] In the nested logit model, the total utility of consumer i choosing flight f_{ij} and currency payment c during decision occasion j is $W_{ij}^{c,f} = U_{ij}^{f} + U_{ij}^{c,f} + \epsilon_{ij}^{f} + \epsilon_{ij}^{c,f}$, where $U_{ij}^{c,f}$ and $\epsilon_{ij}^{c,f}$ are the systematic and stochastic disturbance components of utility specific to the choice combination of flight f_{ij} and currency c (for their definition, see Section 3 in the main manuscript). The maximum total utility $\max_c W_{ij}^{c,f}$ of the currency choices for a consumer i who chose flight f_{ij} during decision occasion j, is assumed to be Gumbel distributed with scale parameter ϕ (Ben-Akiva and Lerman 1985).

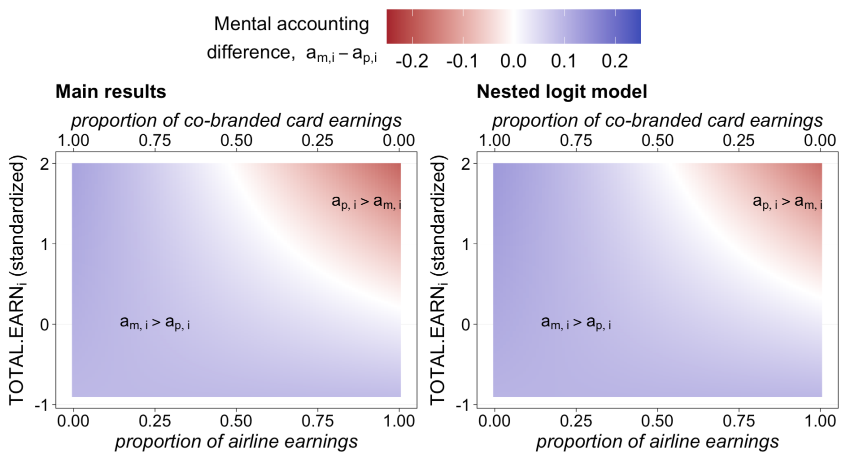


Figure 24 **Effects of point earning source variations on the mean of the mental accounting differences.**

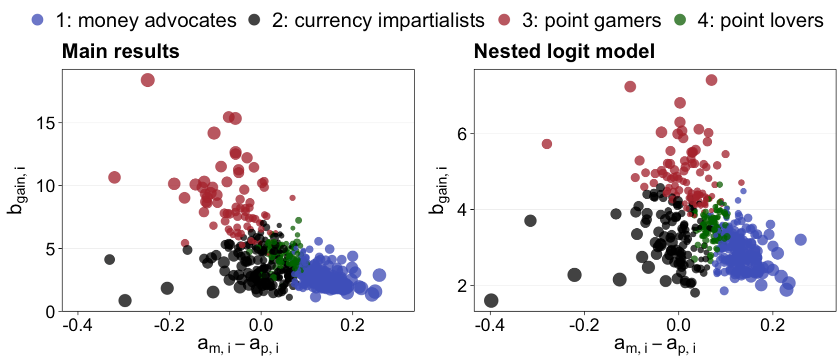


Figure 25 **k-medoids probabilistic segmentation of consumers. The scales for $b_{gain,i}$ are kept different in order to display the segmentation.**

f_{ij}, apart from attractiveness of flight which has been controlled for. The posterior credible interval of the coefficient for flight attractiveness, $\kappa^{ATTRACT} \times \phi$, is positive, which is also consistent with the augmented data since we have simulated a high probability of a consumer choosing a more attractive flight. Finally, Figures 24 and 25 also show that the main results remain robust to the incorporation of flight choice in a nested logit model. In particular, the qualitative conclusions that we draw from them remain unchanged.

References

Ben-Akiva M, Lerman SR (1985) *Discrete Choice Analysis: Theory and application to travel demand* (The MIT Press).

Dudoit S, Fridlyand J (2003) Bagging to improve the accuracy of a clustering procedure. *Bioinformatics* 19(9):1090–1099.

Evgeniou T, Guecioueur A, Prieto R (2022) Uncovering sparsity and heterogeneity in firm-level return predictability using machine learning. *Journal of Financial and Quantitative Analysis* 1–36.

Lahiri K, Gao J (2002) Bayesian analysis of nested logit model by Markov chain Monte Carlo. *Journal of Econometrics* 111:103–133.

Papastamoulis P (2016) label.switching: An R package for dealing with the label switching problem in MCMC outputs. *Journal of Statistical Software, Code Snippets* 69(1):1–24.

Papastamoulis P, Iliopoulos G (2010) An artificial allocations based solution to the label switching problem in Bayesian analysis of mixtures of distributions. *Journal of computational and graphical statistics* 19(2):313–331.

Roels G, Staats BR (2021) OM Forum-people-centric operations: Achievements and future research directions. *Manufacturing & Service Operations Management* 23(4):745–757.

Rousseeuw PJ (1987) Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. *Journal of Computational and Applied Mathematics* 20:53–65.

Train KE (2009) *Discrete Choice Methods with Simulation* (Cambridge University Press), second edition.

Yang Y, Zhang K, Kannan P (2022) Identifying market structure: A deep network representation learning of social engagement. *Journal of Marketing* 86(4):37–56.