TI 2021-077/VIII
Tinbergen Institute Discussion Paper

From Silver to Platinum: the effect of frequent flier tier levels on airline demand

Revision: May 2022

Christiaan Behrens[1]
Gerben de Jong[2]
Jos van Ommeren[3]

1 SEO Amsterdam
2 Hebrew University
3 Vrije Universiteit Amsterdam

From Silver to Platinum: the switching costs created by frequent flier tier levels ⋆

Christiaan Behrens[a,b], Gerben de Jong[a,b,c], Jos van Ommeren[a]

[a] *Vrije Universiteit Amsterdam, School of Business and Economics, De Boelelaan 1105, 1081 HV Amsterdam, The Netherlands*
[b] *SEO Amsterdam Economics, Roeterstraat 29, 1018 WB Amsterdam, The Netherlands*
[c] *Hebrew University, Mount Scopus, Jerusalem 91905, Israel*

Abstract

We estimate the switching costs created by tier levels, one of the main components of frequent flier programs, by exploiting discrete tier thresholds. We first demonstrate that members increase their demand to reach a higher tier level just before the end of the calendar year when tier levels are determined, but do not manipulate demand in earlier months. This allows for a novel fuzzy discontinuity approach to identify causal demand effects of higher tier levels, from which we derive the switching costs. While the lowest level creates only negligible switching costs, the switching costs associated with the highest tier level are in the range of 30 - 41% of the price of a ticket. These results are consistent with the use of tier levels to induce loyalty from high-demand members who are not affected by free flight awards.

Keywords: switching costs, frequent flier programs, customer loyalty, fuzzy regression discontinuity, manipulated running variables

⋆ *Acknowledgements:* We would like to thank an anonymous frequent flier program for providing the data. We thank Timo van Ommeren for research assistance. Furthermore, we thank audiences of several seminars/conferences for useful comments (VU University, Amsterdam; SEO Amsterdam Economics; ITEA, Barcelona; ATRS, Antwerp; ATRS, Amsterdam; SMU, Singapore; Hebrew University, Jerusalem). Jos van Ommeren is a fellow of the Tinbergen Institute.

Email addresses: `c.l.behrens@vu.nl` (Christiaan Behrens), `g2.de.jong@vu.nl` (Gerben de Jong), `jos.van.ommeren@vu.nl` (Jos van Ommeren)

1. Introduction

Airline frequent flier programs are among the most well-known customer loyalty plans in the world. These plans reward members with free flights, upgrades, access to lounges and other services with a value of billions of dollars (Basso et al., 2009). Airlines use frequent flier programs to induce customer loyalty by creating switching costs for members (Klemperer, 1987, 1995). These programs also exploit an agency relationship between employers who pay for airline tickets and employees who book the travel and collect the program rewards (Borenstein, 1996; Basso et al., 2009), as well as benefit from tax advantages since in most countries these rewards are not taxed as income (Levine, 1987; Basso et al., 2016). Although empirical studies confirm that frequent flier programs have economically significant impacts on competition and market prices (Lederman, 2007, 2008),[1] there is remarkable little empirical evidence on how these programs affect members' behavior.

Frequent flier programs reward loyalty by offering benefits, which are a convex function of mileage. If benefits were exactly proportional to mileage they would not lock in customers, as noted by a range of authors including Levine (1987); Banerjee and Summers (1987); Klemperer (1995); Borenstein (1996). One way of offering convex benefits is by giving a free flight after the customer has flown a certain mileage or number of flights, which provides an incentive to concentrate purchases at a single airline. However, offering free flights as a linear function of mileage flown does not create substantial switching costs for high-demand members, as these customers may receive free flights in multiple competing programs (see, e.g., Dowling and Uncles, 1997).[2] In addition, the marginal benefit of a free flight is likely diminishing, which may be another reason why offering free flights for high-demand members does not induce switching costs.[3]

To create switching costs for high-demand members, frequent flier pro-

[1] It is also documented that airlines care a lot about their frequent fliers, see e.g. Gans et al. (2021) for a recent analysis using tweets.

[2] To show this formally assume that customers make annually n flights and discount the future at a rate $1-\delta$. If a loyalty program offers one free flight after x flights, then splitting demand over two airlines decreases the benefit of the program by $(\delta^{x/n} - \delta^{2x/n})/x$. Hence, the benefit of concentrating all purchases in one program is an exponentially decreasing function of the number of flights, n.

[3] See also Hartmann and Viard (2008) for an analysis showing that pure frequency reward programs create only economically relevant switching costs for low-demand members.

grams offer additional benefits through tier levels, which are offered to members if they have flown more than a certain discrete annual threshold defined by number of flights or mileage. Tier benefits are increasing in tier level and typically include benefits such as access to airport lounges, free upgrades and preferential check-in treatment.

The current paper is the first study that estimates switching costs induced by tier levels. Our empirical methodology is based on the fundamental idea that switching costs are revealed through changes in members' demand. We are able to identify switching costs by estimating the effect of tier levels on members' demand using several years of microdata on monthly flying behavior in a major international frequent flier program.

It is useful to distinguish between different types of behavioral demand effects of tier levels. It is well known that tier levels aim to induce members to increase their demand in order to obtain a higher tier level, which we will label as the *incentive effect*. This effect has been demonstrated by, among others, Orhun et al. (2022) and Chen and Ovchinnikov (2019). In order to gauge the members' switching costs, we will focus on another type of behavioural demand effect. We will show that tier levels induce elite members to increase their demand to enjoy the benefits associated with a higher level, which we label as a *consumption effect*.

The consumption effect is measured by exploiting several thresholds corresponding to different tier levels. Due to the presence of the incentive effect, estimation of the consumption effect is not straightforward. That is, the incentive effect implies that members *manipulate* their demand around tier thresholds. Hence, members that just obtain a higher tier level are likely to differ from those that just do not qualify. This invalidates standard econometric methods to estimate the consumption effect, such as a regression discontinuity design using exogenous variation in tier membership around the threshold values (e.g., Imbens and Lemieux, 2008; Lee and Lemieux, 2010).

We use an enhanced estimation strategy, which exploits that manipulation of the running variables – annual flights and mileage – is demonstrated to occur at the end of the calendar year, but not earlier. For this reason we will refer to the *end-of-the-year* incentive effect. The key insight is that the consumption effect can then be identified by employing a fuzzy regression discontinuity design using running variables from earlier months.

The end-of-the-year incentive effect is demonstrated using conventional manipulation tests (McCrary, 2008; Frandsen, 2017), which indicate that members strongly increase their demand to get upgraded, especially for higher tier levels. The strongest effects are for *flight* thresholds. This is in line with it being easier to make one additional flight to just pass a flight threshold than to fly a certain number of miles to pass a mileage threshold. Most importantly, manipulation only occurs in December, while for earlier months, manipulation cannot be detected. This result makes sense because members are hardly able to *precisely* predict future demand, and therefore have little incentive to manipulate until the end of the year approaches.

Our estimates of the consumption effect using the fuzzy regression discontinuity approach – where we use whether the running variable exceeds the threshold in November as the intention-to-treat indicator – show that higher tier levels induce members to increase their demand. Again the effects are most evident on *flight* consumption. The average flight demand effect is equal to about 25%. The estimated tier level specific effects are heterogeneous: they depend on the tier level and are generally stronger for higher levels. For example, the highest tier level induces members to increase their demand by 8 flights, while impact of the lower levels are typically too small (between 0.5 - 2.5 flights) to be detected separately at conventional significance levels.

In conclusion, our paper makes two main contributions. First, we show that airlines use tier levels to create switching costs for high demand members, who are not affected by free flight awards, thereby incentivizing these members to increase their consumption with the airlines in the program. Our point estimates suggest that the switching costs induced by tier levels are equivalent to about 30 - 41% of the ticket price for the highest level, 13 - 17% for the middle level, and 3 - 8% for the lowest level. This novel empirical evidence confirms the theoretical ideas in Levine (1987); Banerjee and Summers (1987); Klemperer (1995), among others, suggesting that offering benefits that increase more than proportionally with mileage is crucial to the effectiveness of loyalty programs. Furthermore, we derive these switching costs from the consumption effect, whereas extant empirical work only considers the incentive effect (e.g., Orhun et al., 2022; Chen and Ovchinnikov, 2019). Given the magnitude of the consumption effect, especially for higher tier levels, this represents an important novelty to the literature.

Second, our identification approach using running variables from earlier months in a fuzzy setup can be applied to other contexts. Strategic behavior around discrete thresholds is a pervasive phenomenon in a broad range of economic domains (see, e.g., Urquiola and Verhoogen, 2009; Camacho and Conover, 2011; Sallee and Slemrod, 2012; Kleven and Waseem, 2013; Gerard et al., 2020). In many contexts manipulation strongly varies over time. For instance, students (and teachers) may be more likely to influence exam results at the end of academic periods. Moreover, welfare benefits typically depend on earnings defined for a certain period, thereby inducing time-specific manipulation. The same is true for many of the economic support measures that are implemented throughout the world to help firms dealing with the economic consequences of the Covid-19 pandemic. There are other strategies to tackle the issue of manipulated running variables, including those that bound the treatment effects (e.g., Dong, 2019; Gerard et al., 2020). A clear benefit of exploiting time patterns in manipulation is that the causal effects remain point identified.

The rest of the paper is organized as follows. Section 2 introduces the frequent flier program and the data. Section 3 describes our approach at estimating the incentive and consumption effects. The results on the incentive effect are presented in section 4, and the results on the consumption effect in section 5. Section 6 derives the tier level specific switching costs from our estimates of the consumption effect. Section 7 concludes.

2. Setting and data

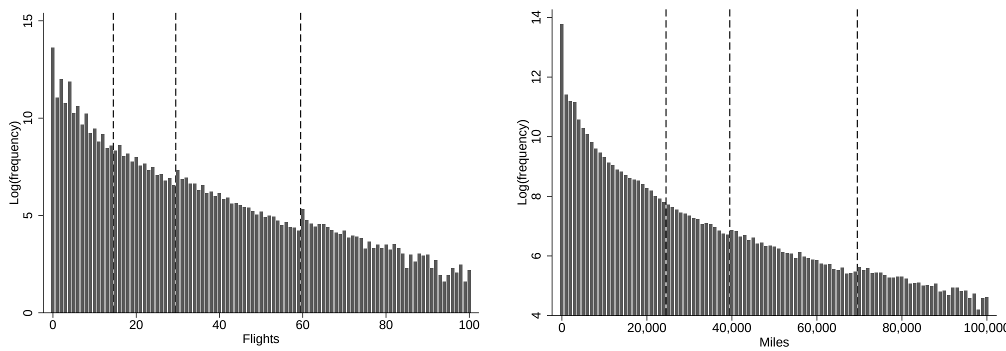
2.1. Basics of the frequent flier program

The frequent flier program that we study represents several major international sponsoring carriers active in different countries. Besides free flight awards, the program offers a tier structure consisting of an introductory level (Bronze) and three elite levels (Silver, Gold, Platinum). The introductory level does not offer any benefit to members except for the possibility to earn free flight awards. The benefits increase with each subsequent tier level. For example, a Platinum member is more likely to get a free upgrade to business class than a Silver member.

In the program - as in most frequent flier programs - members receive a higher tier level (for one calendar year) when on 31 December (or before) their *annual* flights or mileage exceed a certain threshold level.[4] Flights are measured per flight leg, i.e. a return-trip counts as two flights. Mileage is measured as a combination of actual mileage and booking class. Not all

[4]Threshold levels are the same for all countries, except one. We exclude residents from this country.

Figure 1: Distribution of annual flights and miles, thresholds indicated by dashed lines.



(a) Annual flights

(b) Annual miles

flights and miles qualify in the measure that determines tier level. Hence we distinguish between qualifying flights and miles and (all) flights and miles. When elite members do not reach the threshold corresponding to their current elite level by the end of the year, they are downgraded by one level.

2.2. Descriptives full sample

We employ a 20% sample of all members for the years 2013, 2014 and 2015.[5] The data includes the monthly number of flights, mileage and tier level for 485,618 members. About 8% of the members have an elite tier level.

Figure 1 shows the logarithm of the distributions of annual flights and miles. There are several messages in this figure. There is a large difference of occurrence between odd and even numbered flights, which creates a see-saw in the flight distribution, so the flight distribution is not smooth. We will take this into account when focusing on the incentive effect in terms of flights, as manipulation tests are based on smoothness assumptions. For mileage, the smoothness assumption seems reasonable. Both distributions appear monotonically decreasing, and ignoring the notches around the thresholds (indicated by dashed vertical lines), the logarithm of these frequencies are locally approximately linear, another feature which we will exploit when estimating

[5]We exclude special members such as members with Platinum for Life (received by having Platinum for 10 consecutive years) and members of Parliament who are invited to become a member and who receive automatically an elite tier level. For these members, tier level is not related to previous-year flight behavior within the program.

Table 1: Qualifying through flights or mileage

		Qualified through:			
	Nr. of observations	Flights	Miles	Both	Down
Silver	43,627	0.440	0.306	0.117	0.136
Gold	22,149	0.296	0.542	0.087	0.075
Platinum	10,678	0.115	0.839	0.046	
Total	76,454				

the incentive effect. It is also evident that a large share of members do not make any flight.

Table 1 shows that we have information for about 76,000 elite members, of which the majority, about 44,000 have Silver. This table also gives information on whether whether they obtained their tier level based on number of flights, mileage, or, alternatively, on both, or on their previous year status (i.e., taking advantage of the rule that members are never downgraded more than one tier level). The majority of elite members have qualified based on flights or on mileage, but rarely on both. Particularly for Silver, the share of members that qualified based on previous year status is non-negligible. We further note that for Platinum essentially all members qualify based on miles, whereas qualifying based on flights is rare. Consequently, the miles threshold becomes increasingly important for higher tier levels, whereas the flights threshold is numerically important only for lower tier levels.

The program is structured such that the incentive and consumption effects are unlikely the same with regard to moving upwards (e.g. from Silver to Gold) or downwards (e.g. from Gold to Silver).[6] Because the frequencies of flights and mileage are *strongly* decreasing, we have relatively few observations (just right of the threshold) to estimate effects for members that are moving down in the program resulting in non-informative estimates with large standard errors.[7] Consequently, we will only focus on the effect for members who are upgraded if they exceed a threshold.

[6]For instance, the program allows elite members to carry over qualifying miles to the next year *except* when they are downgraded. This rule increase the benefits of passing the tier threshold – the incentive effect – for elite members who are potentially downgraded. Moreover, members who have just lost a tier level keep their benefits until April. Hence, the increase in demand – the consumption effect – for upgraded members must be about one third larger than the induced drop in demand for downgraded members.

[7]The estimation results of the downgraded analysis are available upon request.

Table 2: Descriptives of subsamples around six tier thresholds.

	Mileage threshold		Flights threshold	
	Mean	Std.dev.	Mean	Std.dev.
Silver				
Flights	8.08	5.97	8.31	4.02
Qualifying flights	8.05	5.85	8.38	3.76
Miles	14,443	5,441	8,691	7,358
Qualifying miles per flight	3,050	1,891	1,194	1,042
Nr. of observations	74,961		149,122	
Nr. of individuals	62,600		111,704	
Gold				
Flights	14.28	9.92	21.10	6.88
Qualifying flights	14.23	9.73	21.11	6.36
Miles	26,144	7,780	19,565	13,555
Qualifying miles per flight	3,445	2,102	1,135	896
Nr. of observations	21,228		28,197	
Nr. of individuals	17,429		21,028	
Platinum				
Flights	20.60	13.26	39.38	10.93
Qualifying flights	20.51	13.11	39.20	10.45
Miles	38,388	12,483	29,986	19,545
Qualifying miles per flight	3,700	2,135	974	758
Nr. of observations	15,159		9,694	
Nr. of individuals	11,095		6,621	

Note(s): Subsamples defined based on the following bandwidths: Silver mileage (+/- 15,000 miles), Gold mileage (+/- 15,000 miles), Platinum mileage (+/- 30,000 miles), Silver flights (+/- 10 flights), Gold flights (+/- 15 flights), Platinum flights (+/- 30 flights).

2.3. Descriptives subsamples around tier thresholds

To estimate the incentive and consumption effects, we will apply notch as well as regression discontinuity analyses that are local estimators using subsamples. More specifically, we will employ subsamples of members with a number of flights or mileage close to thresholds. As we will have six thresholds, we focus on six different subsamples around tier thresholds for a given bandwidth.

Table 2 shows descriptives for the different subsamples around six thresholds. For example, we have almost about 15,000 observations in the subsample around the Platinum mileage threshold. Members around this threshold fly yearly about 38,000 miles, and make about 20 flights, i.e. about 10 return trip flights. The average mileage per flight is almost 4,000 miles. Hence, they make rather long/expensive flights. This also holds for the subsamples of members around the Gold and Silver mileage thresholds. In contrast, members around the Platinum flight threshold make almost twice as many flights, but these flights are very short/inexpensive.

In general, members around the flight thresholds make more, but shorter, flights, compared with members around the corresponding mileage thresholds. The Silver flights threshold is an exception as members around this threshold make approximately the same number of flights as those around the Silver mileage threshold (i.e., 8 of trips, hence about 4 return trips).

3. Methodology

Our empirical approach exploits the presence of thresholds to identify the (end-of-the-year) incentive and consumption effects of tier levels. The incentive effect is identified based on manipulation tests which identify *notches* in the frequency distributions of annual qualifying flights and mileage. The consumption effect is identified by a regression discontinuity design using qualifying flights and mileages (in the previous year) as running variables and flights and mileages (in the current year) as the outcomes. A fundamental issue in identifying the consumption effect is that the presence of the incentive effect gives rise to manipulated running variables which invalidates standard regression discontinuity estimates (McCrary, 2008). We solve this issue by using a fuzzy approach which exploits that manipulation occurs in the last month of the calendar year, but not before.

3.1. Identification of the (end-of-the-year) incentive effect

Identification of the incentive effect rests on the assumption that the distributions of qualifying flights and mileage would be smooth in the absence of tier thresholds. The thresholds induce members to manipulate their demand by making more flights or accumulating additional mileage in order to obtain a higher tier level, which creates notches in the frequency distributions.

Note that members' incentive to manipulate increases monotonically over time within the calendar year. Up to the last month(s) of the calendar year, members have little or no incentive to manipulate demand, as they are not able to precisely predict their demand for the rest of the year. However, in the last months of the year, particularly in December, members with a value just below the threshold have an incentive to increase their demand to pass the threshold inducing a notch in the distribution (i.e., a negative notch just before the threshold; a positive notch on and after the threshold).

To formalize this, let us denote $F^q_{i,t}$ and $M^q_{i,t}$ as the annual number of qualifying flights and mileage, respectively, of member i in calendar year t.[8] We focus here on the running variables $F^q_{i,t-1,m}$ and $M^q_{i,t-1,m}$, the *cumulative* number of qualifying flights and mileage respectively in month m of year $t-1$ for $m = 1,..,12$ (this notation implies that $F^q_{i,t-1,m} = F^q_{i,t-1}$ and $M^q_{i,t-1,m} = M^q_{i,t-1}$ for $m = 12$). Our interest is in the probability density functions $g_m(\cdot)$ of $F^q_{i,t-1,m}$ and probability density functions $h_m(\cdot)$ of $M^q_{i,t-1,m}$. A manipulation test in this context is a hypothesis on the continuity of the densities $g_m(\cdot)$ and of $h_m(\cdot)$ at the threshold value z for $m = 1,..,12$.

To estimate the *incentive effect*, i.e. to test for the presence of manipulation, we test for manipulation at the end of the calendar year, i.e. in December, so $m = 12$.[9] In case we do find evidence of the incentive effect, we test whether manipulation was already present in earlier months starting with November, so $m = 11$. As we do not find evidence of manipulation up to November for any threshold, we do not test in earlier months. Hence, we will conclude that members never manipulate up to November (i.e., all manipulation occurs in December).

We emphasize here some further issues which are relevant when testing for manipulation in flight and mileage demand. First, flight and mileage distributions are strongly monotonically decreasing, hence, our graphical evidence to identify notches will use the logarithm of these distributions.

Second, members usually make return trip flights — the standard booking unit is two flights — and only occasionally book an odd number of flights per booking (e.g., an one-way flight, or three flights when travelling to multiple destinations). Hence, odd frequencies occur less often than even frequencies, which implies that the annual flight distribution is not smooth *even in the absence of manipulation*. We solve this issue by focusing on flight distributions separately for odd and even frequencies.[10]

[8]Throughout, we use the superscript q to indicate qualifying flights and mileage and omit this superscript to indicate (all) flights and mileage.

[9]This procedure has good power, despite ignoring information from previous months, because the likelihood of manipulation is an increasing function of month, hence the size of the discontinuity is larger in December than in previous months.

[10]This implies that we test for the presence of notches with a width of (at least) two

Third, the number of flights is a *discrete* variable, so manipulating tests that are based on continuous variables, such as McCrary (2008), are less informative. We will therefore apply the Frandsen test that explicitly acknowledges the discrete nature of number of flights (Frandsen, 2017).[11] Similar to the McCrary test, it makes a smoothness approximation of the distribution (around the threshold). We will assume that the distribution is locally linear, which maximises the power of the test.

We also introduce an alternative test, which does *not* assume that the distribution of number of flights is approximately locally linear, as the latter is inconsistent with our observation that the flight distributions are convex monotonically decreasing (as a consequence, the Frandsen test will overreject the null hypothesis of no manipulation in the absence of manipulation). This alternative test allows that the flight distribution is locally convex by assuming that the number of flights is a local discretized version of an exponential distribution with unknown parameter λ. We emphasise *local*, because if the exponential distribution assumption exactly holds away from the threshold, then the McCrary test would be consistent despite the use of a discrete variable (Frandsen, 2017). For details we refer to Appendix C.2.

Fourth, members will change their behaviour around a certain threshold only when passing of this threshold provides a higher tier level. Such a threshold will be labelled as a *relevant* threshold. Frequently, thresholds are not relevant for members. For example, for a member with Platinum, none of the Silver thresholds are relevant. Our analysis will be based on subsamples of members close, i.e. within a certain bandwidth, to relevant thresholds.

3.2. Identification of the consumption effect

We aim to estimate the consumption effect of having a higher tier level, i.e. the causal effect of tier level on flights and mileage for members who otherwise would have kept the same tier level, which is the counterfactual. We observe monthly flight demand for individual members for three consecutive years, but do not have any information about member characteristics (except for country of residence).

---

flights. This is not problematic, because the dominance of return flights implies that the width of the notch, if present, will be at least two flights.

[11]Manipulating tests that are based on continuous variables, such as McCrary (2008), are potentially misleading. Recent studies have developed manipulating tests which acknowledge the discrete nature of running variables (Frandsen, 2017; Cattaneo et al., 2018).

We examine the effect of tier level on number of flights and mileage. Because number of flights is discrete, and we have a large share of zero flights and therefore mileage, our focus will be on count models. Specifically, we use Poisson models, estimated using a *quasi-maximum likelihood* estimator. Despite its name, this estimator does not assume a Poisson distribution, but makes only assumptions about the relationship between the expected outcome and explanatory variables of interest (Cameron and Trivedi, 2005, p.669). The implied log-linear functional form of Poisson models makes sense: the consumption effect is driven by tier level benefits (e.g., free access to the lounge) that are enjoyed *when making a trip*. This implies that the annual benefit of tier level is approximately proportional to flight demand.

For ease of exposition, we now make four simplifications. First, let us suppose that there exists only two tier levels, so we distinguish between an introductory level that offers no benefits and Silver. Second, suppose for now that tier level is only determined at the end of December. Third, assume that the tier level is only determined by mileage. Fourth, suppose that we are only interested in the question whether Silver increases flight demand.

In this case, we aim to estimate the effect of Silver on number of flights using the following specification for the exponential mean function:

$$E(F_{i,t}) = \exp[\alpha + \beta \cdot S_{i,t} + \gamma \cdot X_t], \tag{1}$$

where $E(F_{i,t})$ denotes the expected number of annual flights $F_{i,t}$, and where $S_{i,t}$ is a dummy indicator of having Silver in year t. Hence, $S_{i,t} = 1$ if $M^q_{i,t-1,12} \geq \overline{M}$, where $\overline{M}$ denotes the threshold value for mileage, otherwise $S_{i,t} = 0$. The control variable X_t refers to a year fixed effect. Here, $S_{i,t}$ is a deterministic function of mileage demand in previous year. The above analysis will not provide a causal effect due to omitted variable bias: members with a higher tier level tend to fly more irrespective of their tier level.[12]

To deal with omitted variable bias, we use a regression discontinuity approach which exploits discontinuities in the treatment assignment using qualifying mileage in the previous year as running variables $M^q_{i,t-1,m}$ for $m = 12$. Each member is assigned into a control (lower tier level) and treatment group (higher tier level), depending on whether the running variable exceeds a known threshold.

[12]Inclusion of member fixed effects to control for time-invariant unobserved characteristics does *not* generate consistent estimates of β, because $S_{i,t}$ is a function of $M^q_{i,t-1,12}$. This problem was first noted by Nickell (1981).

The main concern of the above regression discontinuity design is that, for some thresholds, we will find strong evidence of the presence of incentive effects; causing members to the left of the threshold not to be valid as a control group for members to the right of the threshold.[13] We therefore improve on the 'naive' regression discontinuity design by observing that members manipulate their flight behaviour at the end of the year, i.e. in the month of December, and are less likely to do so during an earlier period. In our application, we do not find any evidence of manipulation in November (or for earlier periods) which can be labelled the *intention-to-treat period*.

A fuzzy regression discontinuity design can be interpreted as an instrumental variables approach (Imbens and Lemieux, 2008). Hence, in essence, we will estimate non-linear instrumental variable models using generalized methods of moments (see Cameron and Trivedi, 2005, p.683). To implement this, we introduce the instrument, $Z_{i,t-1,m}$ which defines whether in month m of the year $t-1$, a member passed the mileage threshold to obtain Silver in year t. Hence, we define $Z_{i,t-1,m} = 1$ if $M^q_{i,t-1,m} \geq \overline{M}$, otherwise $Z_{i,t-1,m} = 0$. The coefficient β in Eq. (1) estimated with generalized method of moments identifies the effect of having one higher tier level on flight demand in the current year, taking into account the endogeneity of having a higher level using as an instrument whether a member has passed the threshold in month m of previous year.

An important requirement of the generalized method of moments instrumental variable approach (as in the more common two-stage least squares instrumental variable approach) is that the instrument is strong. In the first stage, we regress $S_{i,t}$ on $Z_{i,t-1,m}$ where we control for the running variable of month m last year using a quadratic specification:

$$\begin{aligned} S_{i,t} = \zeta + \rho \cdot Z_{i,t-1,m} + \eta_1 \cdot M^q_{i,t-1,m} + \eta_2 \cdot M^{q^2}_{i,t-1,m} \\ + \theta \cdot X_t + \varepsilon_{i,t}, \text{where } m < 12. \end{aligned} \tag{2}$$

Let us suppose that m is equal to 11. The coefficient ρ in Eq. (2) can then be interpreted as the increase in the likelihood of treatment for members in year t who passed the relevant threshold in November of previous year,

[13] Consequently, in this case, members self-select themselves into treatment. Manipulation might not invalidate the regression discontinuity approach, but the ability of members to *precisely* control the running variable near a known threshold does (McCrary, 2008; Lee and Lemieux, 2010; Calonico et al., 2014).

compared to those who did not pass the threshold in that month. For a well-specified model, ρ must be between zero and one. For values closer to one, the instrument is stronger. We will see that instruments based on running variables in November are strong for four out of six thresholds with reasonably high values of ρ. We will focus on those four instruments.[14]

When estimating these models, we will take account of a number of further issues. First, in the majority of regression discontinuity designs, there is only one running variable to determine the treatment status (Lee and Lemieux, 2010). We have two running variables: flights and miles in the previous year. To deal with multiple running variables, one usually converts the two running variables into two *separate* fuzzy regression designs with a single running variable (Jacob and Lefgren, 2004; Matsudaira, 2008). We follow this approach.[15] Hence, for example, when we focus on the running variable mileage, we take into account that mileage is not a perfect predictor of tier level, as members may also have qualified through number of flights. As a result, even without correcting for manipulation (i.e., using running variables from earlier months), a fuzzy regression design arises as qualifying mileage (or flights) is not a perfect predictor of tier status next year due to members ability to qualify based on the other indicator.

Second, for both running variables, we have three thresholds. Given our interest in treatment effect heterogeneity (across tier levels) we apply our fuzzy regression discontinuity design to each individual threshold. For example, when using qualifying mileage as a running variable to determine the effect of Silver, we focus on observations of members within a certain window of the mileage Silver threshold. In order to get local average treatment effects, we pool the estimates using inverse variance weighting. As pointed out by Cattaneo et al. (2016) and Bertanha (2020), pooling the estimates from multiple thresholds provides the average local treatment effect weighted by

[14] In contrast, instruments based on running variables before November are typically weak, as the size effect of these instruments becomes small (e.g. 0.1 or even less), and estimates are then not robust to minor changes in specification.

[15] The design is not optimized in this way, which is possible when one combines multiple running variables (Imbens and Wager, 2018). In the current context, there are few advantages to do that, because few individuals are close to the thresholds of both running variables. Moreover, approaches such as Imbens and Wager (2018) are not applicable to fuzzy discontinuity regression designs, which are used by us to deal with the presence of manipulated running variables.

the relative density of members near each threshold. Combining the two running variables — flights and mileage — around three relevant thresholds — Silver, Gold and Platinum — we end up with six separate designs.

Third, the maximum possible mileage bandwidths are 15,000 miles for Silver and Gold, and 30,000 miles for Platinum. Larger bandwidths would overlap with zero or thresholds of other tier levels. The windows for Silver and Gold are quite small, roughly double the average booking distance for a return flight of around 7,000 miles (calculated for members around these two mileage thresholds). The maximum window for Platinum is set much larger, because we have fewer observations, and is about four times the average booking distance of about 7,500 miles of members around that threshold. The maximum flight bandwidths are 10, 15 and 30 flights for Silver, Gold and Platinum, respectively. Pooling the estimates restricts the bandwidth size somewhat, as estimates from individual thresholds may only be pooled if they come from non-overlapping samples. For instance, in term of mileage, the windows become 7,500 for Silver and Gold, and 22,250 for Platinum.

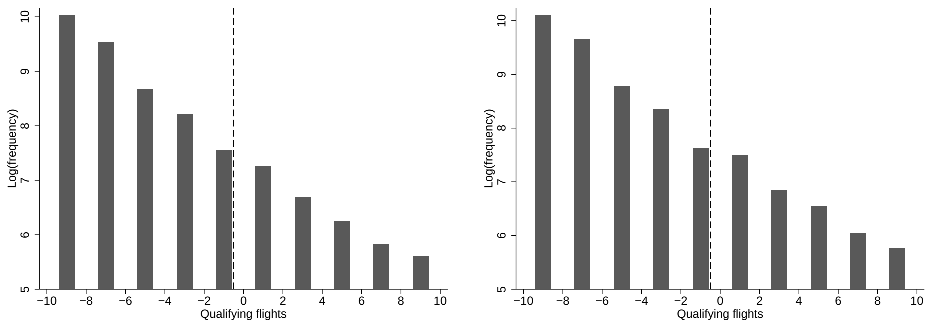
Fourth, not knowing the true functional form of the effects of the running variable, we approximate the functional form by a polynomial function. The non-parametric method of estimating the treatment effect, as introduced by Hahn et al. (2001), involves using (weighted) least-squares regression techniques to estimate the above equation. This requires specifying the order of the polynomial. In fuzzy designs, quadratic polynomials are usually preferred, as a linear specification generates biased results because, the first stage, i.e. the probability of treatment, is poorly predicted.

Fifth, up to now we have assumed that a member who has not qualified in December of previous year will not qualify for a higher tier level until December of the following year. However, a (small) share of members qualify earlier (usually in October of November). We deal with this by indicating tier level membership with a continuous function $S_{i,t} \in [0, 1]$, where $S_{i,t}$ is equal to the proportion of the year that the member receives a higher level. For example, if a member qualifies at October 1, $S_{i,t}$ equals 0.25, as this member is treated for a quarter of year.

4. The presence of incentive effects and the timing of manipulation

This section provides graphical evidence, supported with statistical tests, of *i)* the presence of incentive effects, i.e. whether members manipulate in December, and *ii)* the timing of manipulation, i.e. whether members also

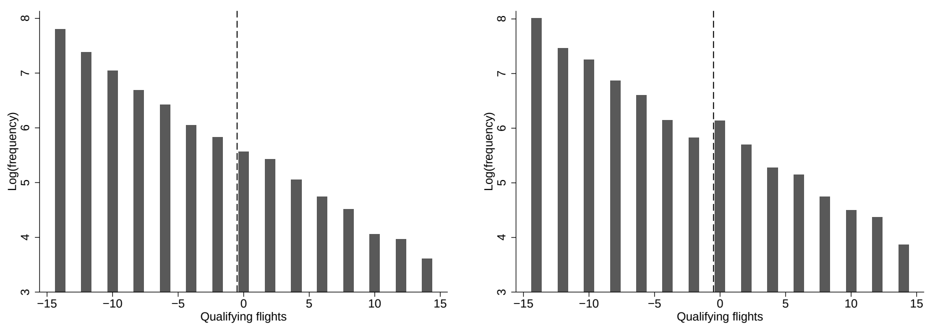
Figure 2: Distribution of qualifying flights around Silver thresholds, even numbers.



(a) November

(b) December

Figure 3: Distribution of qualifying flights around Gold thresholds, even numbers.



(a) November

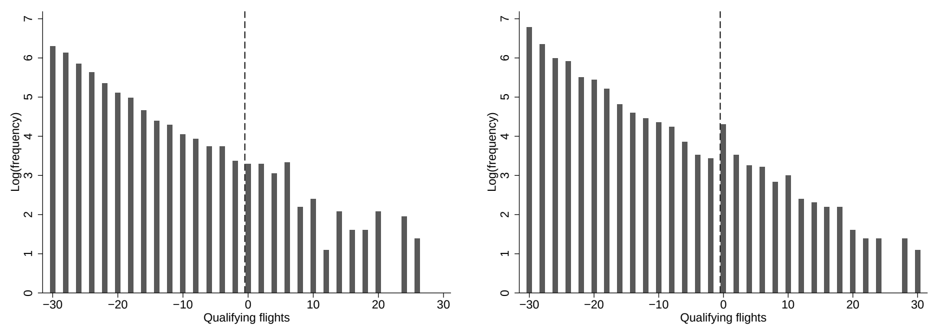
(b) December

manipulate in earlier months. We focus on notches around all three thresholds of both the flights and mileage distributions. The flight distributions for even and odd number of flights are considered separately. As even and odd distributions turn out to be almost identical, we show here the figures of the even distributions (which are based on more observations), whereas for the odd distributions we refer to Appendix A.[16]

Figures 2, 3 and 4 show that the flight distributions exhibit substantial notches around the Gold and Platinum thresholds in December (right panels). The frequency at the Platinum threshold is about one logarithm point higher

[16]The Silver threshold is an odd number, whereas the other two flight thresholds are even. Hence, for Silver, the threshold is in between two even numbers.

Figure 4: Distribution of qualifying flights around Platinum thresholds, even numbers.



(a) November

(b) December

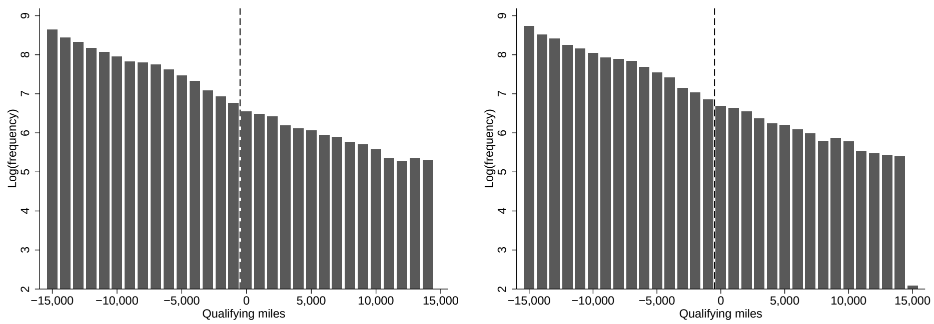
than the frequency just before this threshold. Consequently, members are 70% more likely to just exceed the minimum threshold of the Platinum level. Around the Gold threshold it is about half a logarithm point. Around the Silver threshold there also seems to be a notch in December, although it is substantially smaller. These results make sense. The additional benefits of Gold and Platinum are higher and, hence, for those who make many flights it is more economical to manipulate. Moreover, one expect that a substantial share of those that make many flights may easily make one more flight.

The graphical evidence on manipulation around the flight thresholds is supported by statistical tests, see Appendix B. We apply several locally linear Frandsen tests for odd and even values separately and combine both test outcomes into a single test using a Bonferroni correction.[17] For all thresholds the p-value is < 0.01.

We arrive at the same conclusion, when we apply the Frandsen test where we already allow for non-linear curvature, i.e. we make the assumption that the number of flights comes from a discretized version of an exponential distribution, where one has to estimate the λ, see Appendix C.1. We also

[17]The Frandsen test requires the researcher to specify, a priori, to what extent the running variable distribution differs from a *linear* distribution using a parameter k. When $k = 0$, the distribution is assumed to be locally linear. When $k > 0$, the test allows for non-linear curvature, but has less power to detect manipulation. We apply the Frandsen (2017) test using $k = 0$, in order to maximise the power of the test. Given that the even and odd samples are independent, we apply the Bonferroni correction: $p = 2 * min(p_{even}, p_{odd})$, where p refers to p-values.

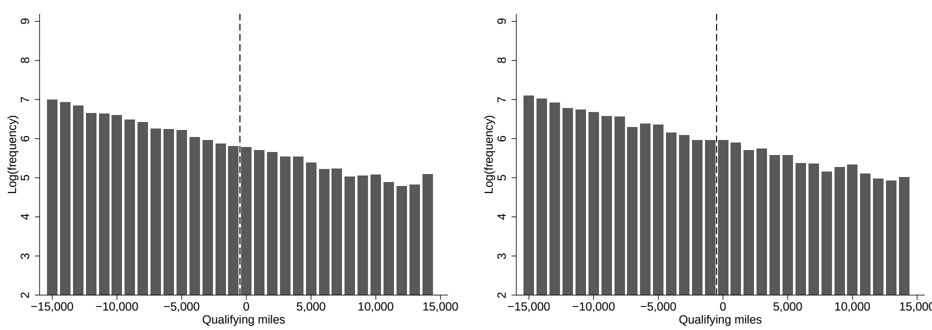
Figure 5: Distribution of qualifying miles around Silver thresholds.



(a) November

(b) December

Figure 6: Distribution of qualifying miles around Gold thresholds.



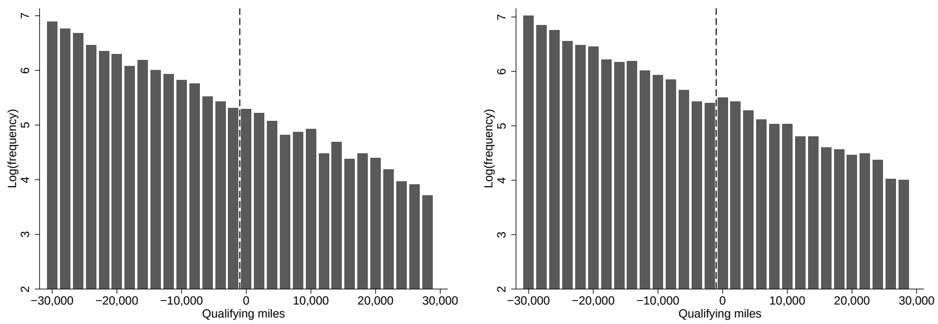
(a) November

(b) December

introduce an alternative test, which makes the same assumption, but where one does not have to estimate λ, see Appendix C.2. Again, we arrive at the same conclusion. Hence, in line with the graphical evidence, the notches at Silver, Gold and Platinum flight thresholds are significant at conventional significance levels under a variety of assumptions.

For the mileage distributions, as depicted in Figures 5, 6 and 7, the notches are again most clearly visible around the Gold and Platinum thresholds in December (right panels), although less pronounced than around the flight thresholds. For example, the notch at the Gold threshold is only about 0.15 logarithm point. For Silver, there is no visible notch. The finding that manipulating is absent for members around the Silver mileage threshold is in line with the idea that it is very difficult for these members to precisely manipulate mileage, and there are likely fewer opportunities to manipulate

Figure 7: Distribution of qualifying miles around Platinum thresholds.



(a) November

(b) December

as members around this threshold make just a few flights.

Again the graphical evidence is supported by the accompanying tests. We now use the McCrary test, which is equal to 0.02 (p-value > 0.05) for Silver, and respectively 0.16 (p-value < 0.05) and 0.32 (p-value < 0.05) for Gold and Platinum (all in December). Hence, the notches at Gold and Platinum are significant at conventional significance levels, whereas the notch at the Silver threshold is not statistically significant. In conclusion, only for the mileage Silver threshold one is allowed to use miles information for December as a running variable to identify the consumption effect.

To obtain information about the timing of manipulation, we now focus on the presence of notches in November. In contrast to the results for December, there is no evidence of manipulation in November *at any of the thresholds* (this also holds for earlier months). This result is supported by graphical evidence, see Figures 2 - 7 (left panels), as well as by statistical tests showing p-values > 0.05 for the thresholds in November (see Appendix B and C).[18] Consequently, manipulating may occur in December, but never before.

The main implication is that to identify the consumption effect for the

[18]The only exception is the alternative test for flights around the Silver threshold, see Appendix C.2, which suggests (in contrast to the corresponding Frandsen tests) that there is manipulation even in November. At the same time, the test indicates that the amount of manipulation is negligible (i.e. at the threshold, the frequency is only slightly larger than its theoretical value given no manipulation). We have nearly 8,000 observations near this threshold, so even a small deviation from the exponential distribution assumption is likely to induce one to falsely reject the null hypothesis of no manipulation.

Gold and Platinum thresholds, one may use a fuzzy regression discontinuity design using the number of qualifying flights as well as qualifying mileage in *November* as running variables. The same fuzzy design can be used to identify the consumption effect at the Silver flights threshold. At the Silver mileage threshold, as concluded earlier, one may simply use the qualifying mileage in *December*, because there is no evidence of manipulation at this threshold.

5. Consumption effect

5.1. Discrete jump in probability of being treated

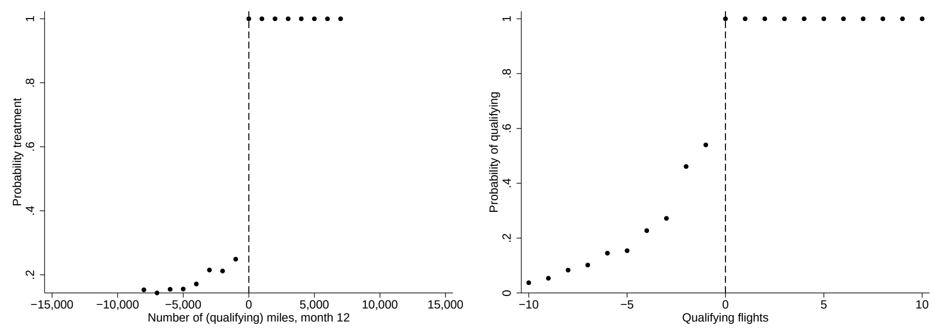
To identify the consumption effect, the fuzzy regression discontinuity design requires the presence of a substantial discrete jump ("strong instrument") of the probability of being treated, i.e. receiving a higher tier level, around the threshold values. We will examine this by providing graphical evidence.

Figure 8 shows that there is a substantial discrete jump in the probability of being treated at both Silver thresholds (i.e., the mileage threshold *in December* and the flights threshold *in November*).[19] For example, members who are just below the mileage threshold in December are about 75 percentage points less likely of being treated than those at or just over the threshold. Members who were just below this threshold and still got treated in the next year qualified through the other indicator (e.g., members who make relatively short flights may end up just below the mileage threshold but still qualify through their number of flights). For members who are one single flight below the flights threshold in November, the discrete jump is approximately equal to 40 percentage points. This discrete jump is substantially smaller as, in addition to qualifying on the other indicator (i.e., mileage), members can also still qualify for treatment based on flights in December.

In Figures 9 and 10, it is shown that there are also substantial discrete jumps for the Gold and Platinum mileage thresholds *in November*. For the Gold mileage threshold the jump is equal to about 50 percentage points. For the Platinum mileage threshold the discrete jump is also substantial, at almost 40 percentage points. For flights, the jumps at both the Gold

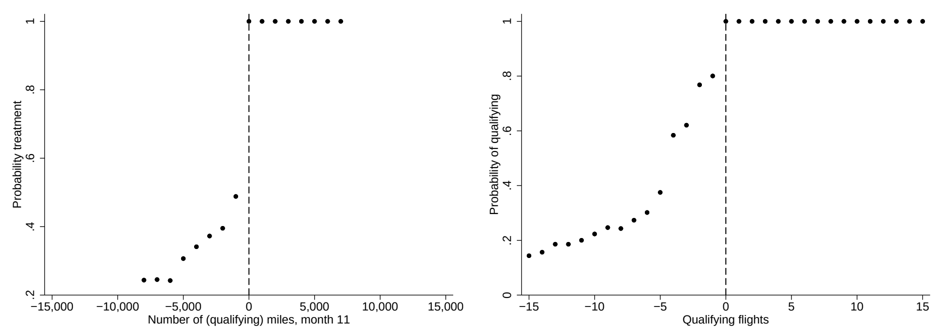
[19] Because we use running variables from November to address manipulation at the Silver flight threshold, November is here the relevant month to check for a discrete jump in the probability of being treated. We repeat this for the Gold and Platinum thresholds.

Figure 8: Probability of qualifying for Silver around thresholds.



(a) Miles, December (b) Flights, November

Figure 9: Probability of qualifying for Gold around thresholds.



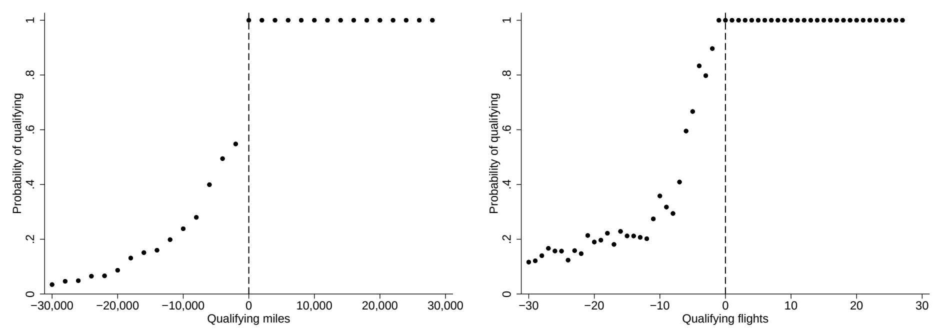
(a) Miles, November (b) Flights, November

and Platinum threshold are much smaller. At the Gold flights threshold the jump is less than 20 percentage points.[20] For the Platinum flights threshold the jump is absent. This implies that for the Gold and Platinum flights thresholds, we do not have a convincing identification strategy. This is not problematic as only a small share of members qualify for Gold and Platinum based on flights (see Table 1).

These figures imply another important message. The probability of being treated is a monotonically increasing but convex function in the running variables. Hence, in the fuzzy discontinuity regression, controlling for the

[20]The threshold value is an even number, so the discrete jump is smaller than suggested by the figure, as even numbers occur more frequently than odd numbers.

Figure 10: Probability of qualifying for Platinum around thresholds.



(a) Miles, November (b) Flights, November

running variable in a quadratic way is adequate to capture the curvature, while when the window size is strongly reduced (e.g., 7,500 miles and 5 flights) controlling for a linear function may give a (slight) underestimate, but not an overestimate.

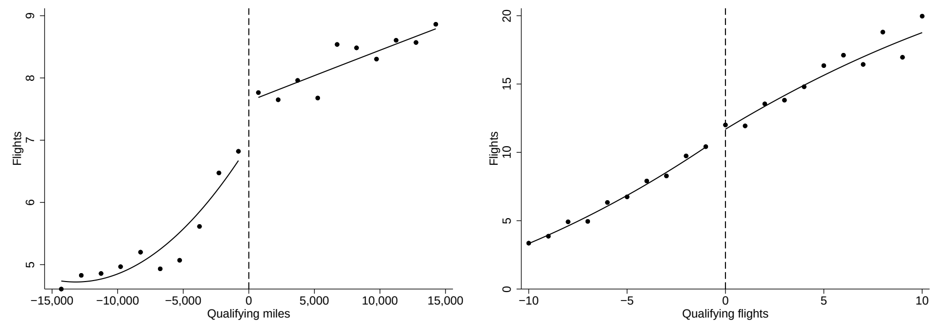
5.2. Graphical evidence of consumption effect

We now provide graphical evidence of a discrete jump in the dependent variables (mileage and flights) at the four thresholds for which we have a convincing identification strategy: the Gold and Platinum mileage thresholds and for Silver both thresholds. It is important to note that the figures in this section are only suggestive as the shown discontinuities illustrate the intention-to-treat effects. These effects are stark underestimates, because as shown in previous subsection, at all thresholds - most notably Gold and Platinum mileage thresholds - there are large shares of non-compliers (i.e., members at the left of the threshold who eventually will get treated), while there are no never-takers (i.e., members at the right of the threshold who are not treated).[21]

Figure 11 indicates that there is a moderate increase in the number of

[21]To phrase it differently employing classical econometric theory language, these figures demonstrate the "reduced form" relationship between the dependent variables of interest and the relevant running variables, which is only partially informative about the size of the "structural form" effects we are interested in. If there is no discrete jump in the dependent variable at the threshold, then the "structural form" effect of tier level, which will be estimated with the fuzzy regression discontinuity design, will be absent.

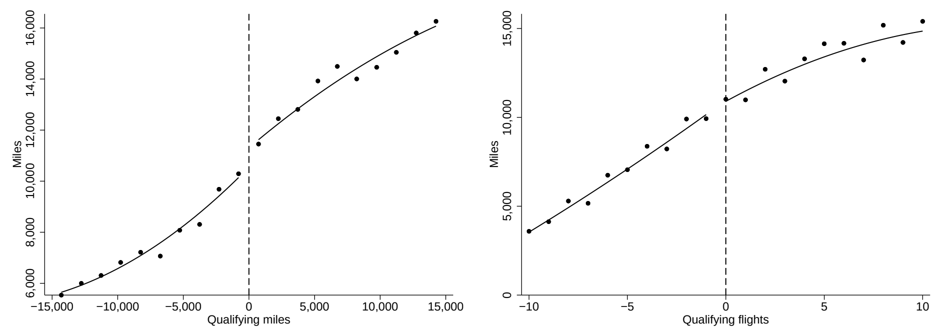
Figure 11: Flights around Silver thresholds.



(a) Qualifying on miles (December)

(b) Qualifying on flights (November)

Figure 12: Mileage around Silver thresholds.



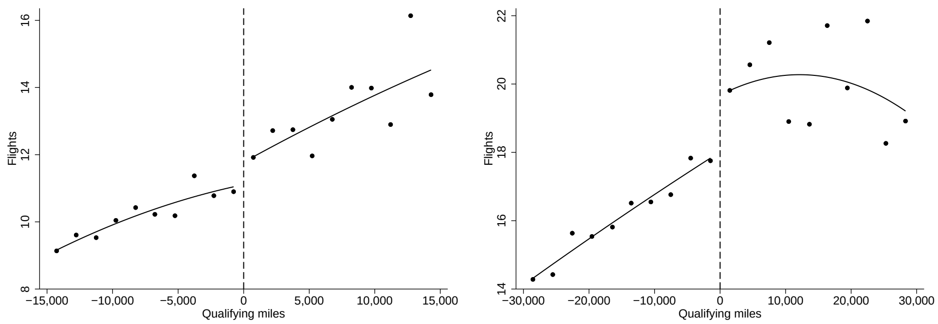
(a) Qualifying on miles (December)

(b) Qualifying on flights (November)

flights at the Silver mileage threshold, but virtually none at the Silver flight threshold. The suggested increase is about 0.4 flights at the mileage threshold. Hence, combined with information about the discrete jump in the probability of treatment of about 0.75 (Figure 8a), the suggestive causal effect on flights of receiving Silver is estimated to be about $0.4/0.75 = 0.5$ flights. Figure 12 shows that there appears to be no increase in the consumption of mileage for those that obtained Silver – neither at the mileage, nor at the flight threshold.

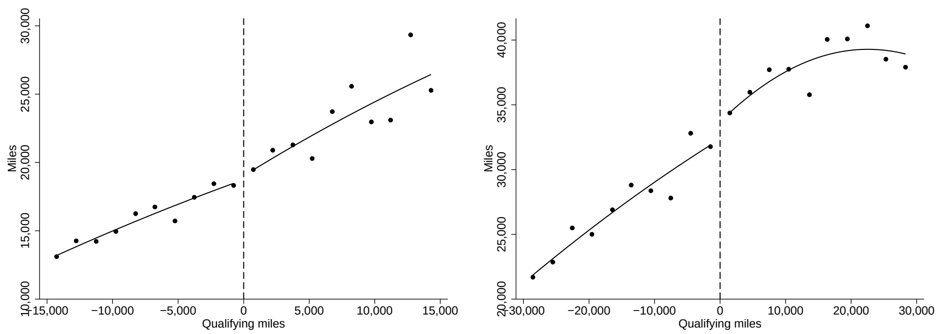
Now we consider Figure 13, showing the flight consumption effects at the Gold and Platinum mileage thresholds. There is a clearly visible discrete increase in the consumption of flights around both these thresholds. For instance, at the Platinum threshold there is an increase of about two flights. Recall that the increased probability of qualifying for platinum (based on

Figure 13: Flights around Gold/Platinum thresholds.



(a) Gold (b) Platinum

Figure 14: Mileage around Gold/Platinum thresholds.



(a) Gold (b) Platinum

running variables in November) is roughly 0.4 (see Figure 10a), hence this implies that the consumption effect of Platinum is about $2/0.4 = 5$ flights.[22] On the other hand, as shown in Figure 14, the increase in mileage demand at the Platinum and especially the Gold threshold seems negligible.

[22] To some readers the figure might suggest that the increase equals about 30% around the threshold, but we will see later that this is a gross underestimate. The intuition is that a large share of the members that are on the left side of the threshold will eventually be treated in December and, hence, the level of flights just at the left side of the threshold is an overestimate of the true flight demand of those that are not treated.

Table 3: Consumption effect at miles thresholds.

	Pooled Poisson		Tier level-specific					
			Silver		Gold		Platinum	
	Overall	Gold-Plat.	linear	Poisson	linear	Poisson	linear	Poisson
Flights	0.246**	0.456***	0.460	0.0760	2.484	0.358	7.920**	0.534**
	(0.109)	(0.163)	(0.646)	(0.147)	(1.849)	(0.245)	(3.858)	(0.219)
Mileage	0.058	0.110	246.4	0.0330	468.8	0.00129	4752.0	0.192
	(0.092)	(0.161)	(900.1)	(0.113)	(2826.9)	(0.246)	(5132.8)	(0.214)
Nr. of obs.	28,251	10,935	17,316		5,237		5,698	

Note(s): Fuzzy regression discontinuity estimates of the consumption effect at the miles thresholds. Silver threshold is based on running variable in December of the previous year, while Gold and Platinum thresholds are based on running variables in November. Bandwidths are set at 7,500 miles for Silver and Gold, and at 22,500 miles for Platinum. Robust standard errors are given within parentheses. ***p<0.01; **p<0.05; *p<0.1

5.3. Main analysis consumption effect

We now present the main estimation results using the fuzzy regression discontinuity design for the consumption effect around the mileage and flights thresholds.[23] For the mileage thresholds we present pooled estimates and tier level-specific estimates. For the flights thresholds we only present the estimates for Silver, as we do not have a convincing identification strategy for the Gold and Platinum flights thresholds.

Table 3 presents the estimates based on the mileage thresholds. For each tier level, two estimates are reported: linear estimates based on a two-stage least squares approach, and Poisson estimates estimated by generalized method of moments. The pooled Poisson estimates are obtained by an inverse variance weighted average of the tier level-specific Poisson estimates.[24] These pooled estimates are estimated over all tier levels (in the first column), as well as only over the Gold and Platinum estimates (in the second column).

The first row reports the flight consumption effects, that is, the increase in the demand for flights induced by obtaining a higher tier level. The pooled

[23]First-stage estimates can be found in Appendix D; full second-stage estimation results are provided in Appendix E.

[24]Direct pooling of the Poisson estimates makes sense, as these estimates have a "percentage" interpretation. The pooling method puts greater weights on lower tier levels (i.e., Silver and, to a lesser extent, Gold), because standard errors tend to be smaller for lower tier levels due to larger number of observations. If anything, this puts downward pressure on our pooled estimates given that the effects are increasing in tier level.

Table 4: Consumption effect at Silver flight threshold.

	Silver	
	linear	Poisson
Flights	0.0171	0.0753
	(0.782)	(0.130)
Mileage	-856.1	-0.0294
	(1004.8)	(0.252)
Nr. of observations	87,483	

Note(s): Fuzzy regression discontinuity estimates of the consumption effect at the Silver flight threshold. Running variables based on November of the previous year. Bandwidth is set at 10 flights. Robust standard errors are given within parentheses. ***p<0.01; **p<0.05; *p<0.1

estimates are positive and statistically significant. The overall effect (about 25%) is substantially lower than the (pooled) Gold-Platinum effect (about 45%), reflecting the large weight put on the almost zero effect of the Silver tier threshold.

In line with the graphical evidence, all estimated tier level-specific flights effects are positive, although only the Platinum effect is statistically significant on its own. The effects of lower tier levels appear too small to be detected separately at conventional significance levels. The point estimates increase from about 0.5 flight or 8% for Silver, to 2.5 flights or 36% for Gold, and almost 8 flights or 53% for Platinum.

The second row reports the consumption effects in terms of mileage, i.e. the additional demand for mileage caused by obtaining a higher tier level. All estimates (pooled and tier level-specific) are again positive, although none are statistically significant. Point estimates vary between about 250 miles (or 3%) for Silver, and 4,750 miles (or 20%) for Platinum, with pooled estimates equal to 6% for the overall effect and 11% for the two higher tier levels.

Table 4 presents the consumption effect estimates based on the Silver flights threshold. Consistent with our finding for the Silver mileage threshold, the estimates of the consumption effect at the Silver flights threshold are not statistically significant at conventional levels. The Poisson point estimate (7.5%), however, is almost exactly equal to the one obtained from the Silver mileage threshold (7.6%).[25]

[25]In a sensitivity analysis, we have clustered the standard errors by the digit of the

Together, these results have two main implications. First, the results are in line with members increasing their demand after obtaining a higher tier level. The estimates of the tier-level specific effects are generally positive and the pooled estimates on the flight consumption effects estimates at the mileage thresholds, which retain a large sample size and therefore accuracy, are also statistically significant. The results therefore seem to suggest that member increase their consumption of flights more strongly than their consumption of mileage after obtaining a higher tier level. This implies then that higher tier levels induce members to make more, but shorter, flights (within the program).[26] However, the confidence intervals of the consumption effect of mileage and flights largely overlap, so it is also plausible that these consumption effects are similar.

Second, as evidenced by the estimates at the mileage thresholds where we can consider multiple tier levels, the consumption effects seem to be increasing in tier levels. This accords well with the theoretical idea that the award structure in loyalty programs should be non-linear to create switching costs. As mentioned in the introduction, the effectiveness of the frequency rewards decreases for members at the upper end of the demand range. Hence, airlines use the tier structure to offer additional and increasing benefits through tier levels.[27] These additional benefits effectively ensure the loyalty of those that fly the most.

5.4. Sensitivity analysis and placebo tests

Figures 15 - 16 provide graphical evidence of the robustness of the consumption effect estimates to bandwidth selection. We present here the pooled Poisson estimates at the mileage threshold.[28]

---

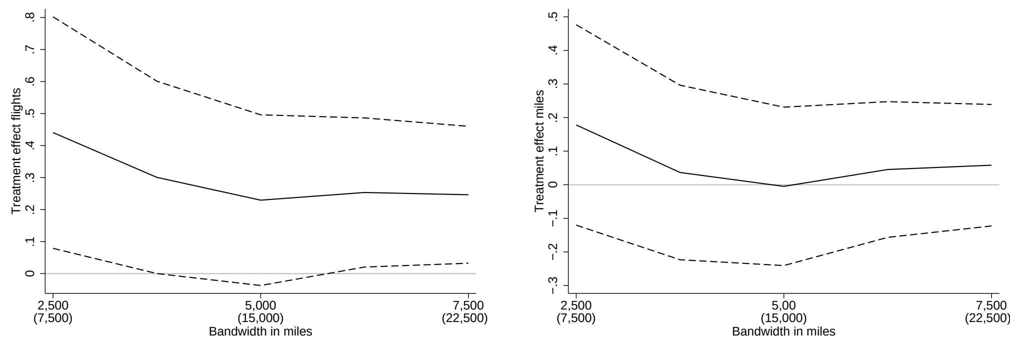
running variable, which takes into account that the running variable is discrete. We find then very similar, statistically insignificant, results as reported above. For a criticism of this approach see Kolesár and Rothe (2018). We have not explored their enhanced approach further, as it is unlikely that the standard errors become substantially smaller.

[26] An explanation for this could be that members substitute short-haul trips on other transport modes to air travel, or substitute away from low cost carriers (who offer predominantly short-haul flights) after obtaining a higher tier level.

[27] For example, with Silver, members do not have access to a lounge, while Gold members do have access. Platinum members almost always get upgraded when the flight is full.

[28] For larger bandwidths, some of the samples overlap. To avoid this issue we reduce windows by the same proportion.

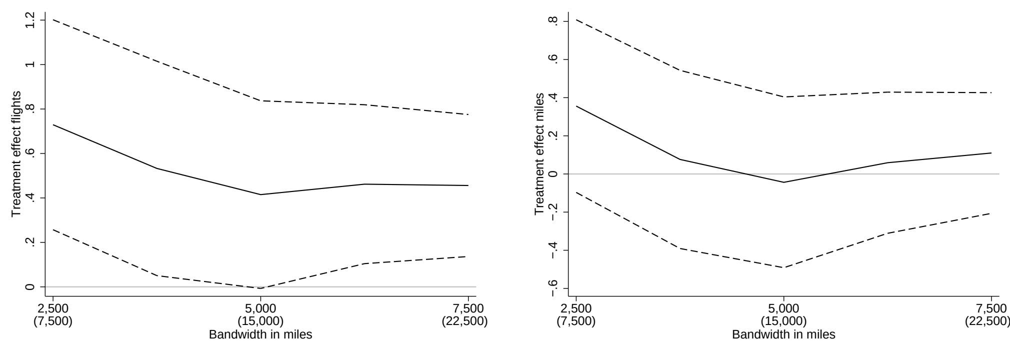
Figure 15: Consumption effect at miles thresholds (pooled estimates).



(a) Flights (b) Mileage

Note(s): Pooled Poisson estimates for different bandwidths; bandwidth for Platinum between brackets.

Figure 16: Consumption effect at miles thresholds (pooled Gold-Platinum).



(a) Flights (b) Mileage

Note(s): Pooled Poisson estimates for different bandwidths; bandwidth for Platinum between brackets.

For the flights consumption effects at the mileage threshold, the bandwidth plots show rather robust patterns with positive estimates ranging from 25 - 45% for the overall pooled effect (Figure 15) and 45 - 75% for the pooled effect of the two higher tier levels (Figure 16). Pooled estimates are statistically significantly different from zero for all window sizes, except for the window size in the middle of the plot (i.e., 5,000 for Silver and Gold, and 15,000 for Platinum). The (pooled) mileage consumption effects at the flights threshold are not statistically significant for any window size.

The tier level-specific consumption effects at the mileage thresholds are also rather stable over the bandwidth range (results can be received open request). For instance, the point estimate for the flights consumption effect at the Gold mileage threshold is equal to approximately 35% (i.e., our baseline estimate) over the full range of window sizes, although the confidence interval includes zero at any chosen window size. The only less stable effect is the consumption effect at the Silver flights threshold. Point estimates range from positive to negative, and confidence intervals vary from very wide and including zero at some window sizes to comparatively small and excluding zero at others.

We have also applied a range of placebo tests by re-estimating models while slightly changing the threshold value. For example, at the mileage thresholds, we have increased (as well as decreased) the mileage threshold by only 1000 miles, and by 5000 miles. Placebo estimation results, as reported in Appendix F, show that the consumption effects are fully absent at these placebo thresholds. Hence, we conclude that our results are robust with respect to a range of placebo tests and take this as evidence that our results are indeed driven by the discontinuous treatment shift (i.e., obtaining a higher tier level) around the actual thresholds of the frequent flier program.

5.5. Consumption effect revisited: intensive and extensive margins

In the above analysis, our focus is on the overall effect of tier levels on consumption behavior. Here, we decompose this effect distinguishing between the *extensive margin*, i.e. whether members make at least one flight per year,[29] and the *intensive margin*, i.e. the amount of travel – in terms of miles or flights – conditional on making at least one flight per year. There are two reasons why this decomposition is of interest.

The first reason is statistical: a non-negligible share of members for which we determine the consumption effect (i.e., members who are in the relevant subsamples around the tier thresholds as reported in Table 2), do not make any flight in the year investigated. For Silver, this share is about 25%, for Gold it is about 11% and for Platinum about 9%. The second, economic, reason is that one expects the extensive margin of a higher tier level to be important: members with a elite tier level have strong incentives to make

[29]The extensive margins for the effect on flights and miles are equal, because individuals who make at least one flight also have a positive mileage.

Table 5: Extensive margin at miles threshoold.

	Pooled linear	
	Overall	Gold-Platinum
Extensive margin		
Flights/Mileage	0.058**	0.073*
	(0.024)	(0.038)
Nr. of observations	28,251	10,935

Note(s): Fuzzy regression discontinuity estimates of the extensive margin of the consumption effect at the mileage thresholds, using a dummy indicator of at least one flight in the program as the dependent variable. Only pooled estimates shown. Running variables and bandwidths similar as in main specification (see Table 3). Robust standard errors are given within parentheses. ***$p<0.01$; **$p<0.05$; *$p<0.1$

all their flights within one program. So loyalty programs aim to prevent members from stopping to fly with them, as conditional on flying once, they are likely to make many flights.

To examine this further, we estimate the effect of tier level on the extensive margin, focusing on the effects occurring at the miles thresholds, as for these thresholds we found evidence of the consumption effect. We apply the same fuzzy regression discontinuity design as for the overall consumption effect, the only difference is that the dependent variable is now a dummy indicator, yielding a linear probability setting.

Table 5 reports the pooled estimates for the extensive margin using the miles thresholds. In line with the above considerations, the estimates provide evidence that the extensive margin is relevant. On average, the probability of making at least one flight is increased by about 6 - 7 percentage points for all the higher tier levels. We consider this to be a substantial effect given that the share of members with higher tier levels that stop flying with the airline is around 10 percentage points each year. Consequently, *the share of elite members that stop flying with the airline is reduced by two thirds due to the tier level benefits*. Furthermore, we consider this effect as large because it is plausible that major events in the life of the members (e.g., a change of job, retirement, or even death) completely nullify their (airline) demand. This implies that the incentive whether or not to fly within the program is not affected by the level of tier level for a non-negligible group of members

Table 6: Intensive margin flights and mileage at miles threshold.

	Pooled 2SLS (log)		Pooled Poisson	
	Overall	Gold-Platinum	Overall	Gold-Platinum
Intensive margin				
Flights	0.031	0.110	0.124	0.338**
	(0.069)	(0.136)	(0.106)	(0.164)
Mileage	-0.039	-0.012	-0.043	-0.018
	(0.085)	(0.157)	(0.086)	(0.161)
Nr. of observations	21,514	9,731	21,514	9,731

Note(s): Fuzzy regression discontinuity estimates of the intensive margin of the consumption effect at the mileage thresholds, estimated on the subsample of members with at least one flight in the program. Only pooled estimates shown. Running variables and bandwidths similar as in main specification (see Table 3). Robust standard errors are given within parentheses. ***p<0.01; **p<0.05; *p<0.1

who experience those events – even if the benefits would approach infinity.[30]

For the intensive margin, we apply exactly the same approach as for the overall consumption effect, the only difference is that we deal with a subsample.[31] Because the dependent variable is now strictly positive, we are able to use a log specification in addition to the Poisson approach, allowing for a standard two-stage least squares approach. In Table 6 we report the intensive margin effects (the full results are reported in Appendix G.1). It appears that the coefficient for the intensive margin is somewhat smaller given the Poisson specification and even statistically insignificant given the log specification. Again, the large effects appear to be concentrated at the higher tier levels.

In conclusion, the evidence indicates that the extensive margin of the consumption effect is important at all tier levels. For the higher levels, the share of elite members that stop flying with the airline is reduced by two thirds due to obtaining a higher level. The intensive margin also plays a role, but mostly at the higher tier levels.

[30] In the data there are a considerable number of cases where individuals with a sizeable number of flights in a given year, drop to zero flights in the subsequent year. We believe this provides evidence of the incidence of major life events on their airline demand.

[31] Note that we ignore the selection effect of the subsample, so the effect estimated should be interpreted as a decomposition effect and not as a causal effect.

6. Switching costs created by tier levels

We proceed by quantifying the switching costs induced by frequent flier tier levels. If consumers with an elite tier level switch (fully) to another airline they loose their tier level benefits. Hence, switching costs are equal to the monetary benefits associated with higher tier levels, i.e. the member's willingness-to-pay for tier level. This willingness-to-pay can be derived from our estimates of the consumption effect.

We assume a representative consumer who chooses to make F flights and whose associated (indirect) utility U is a function of the price of flight tickets p, and the binary indicator S, which measures the tier level. So, $F = F(U(p,S))$. We aim to derive the willingness to pay for tier level as a fraction of the ticket price. Consequently, we are interested in $[\frac{\Delta U}{\Delta S}/\frac{\partial U}{\partial p}]/p$. Given the approximation $\frac{\Delta log(F)}{\Delta S} = \frac{\partial log(F)}{\partial U}\frac{\Delta U}{\Delta S}$, it follows that:

$$[\frac{\Delta U}{\Delta S}/\frac{\partial U}{\partial p}]/p = -\frac{\Delta \log(F)}{\Delta S}/\frac{\partial \log(F)}{\partial \log(p)}. \quad (3)$$

Hence, the benefits of a higher tier level relative to the ticket price can be derived given information on $\frac{\Delta \log(F)}{\Delta S}$ and the demand elasticity $\frac{\partial \log(F)}{\partial \log(p)}$. For the latter demand elasticity, we rely on the literature which typically finds values within the -2 and -1.5 interval (see, e.g., Berry and Jia, 2010).[32] We will proceed on the conservative side with a demand elasticity of -2.

Recall that our point estimates of the effect of a higher tier level on the consumption of flights are 0.08, 0.36 and 0.53 for Silver, Gold and Platinum, respectively. The same estimates for the effects on mileage consumption were substantially lower at 0.03, 0.00 and 0.19, respectively. If we use the within-tier level averages of these estimates and plug these into Eq. (3), then the benefit of having Silver is about 3% of the ticket price; the benefit of having Gold is about 13% of the price; whereas the benefit of having Platinum is 41% of the price.

[32] Note that we do not focus on the overall demand elasticity, but on the demand elasticity *for a certain airline*. The high absolute value of this elasticity in the airline market is in line with studies for other markets, e.g. the car market, where consumers easily switch between suppliers of products which offer close substitutes (Train and Winston, 2007). Note that we assume a constant demand elasticity. Especially for members around the higher tier levels it is likely that their demand elasticity is somewhat lower. This would imply that we underestimate the switching costs for higher tier levels and, moreover, that the differences in switching costs between the lower and higher levels is even more pronounced.

This approach may be criticised, because several of these point estimates are not statistically different from zero at conventional significance levels. A more conservative approach is to rely on the pooled estimates, yielding benefits of 8, 18 and 30% of the ticket price for Silver, Gold and Platinum, respectively. An even more conservative approach is to assume that the statistically insignificant tier-level-specific effects are equal to zero. The benefits for Platinum are then equal to 15% of the price, and zero for the other tier levels.[33]

Our consumption effect estimates point at substantial benefits, and hence switching costs. We will now demonstrate that our estimates of the benefits are also consistent with the estimated (end-of-the-year) incentive effects. In line with our findings, we assume the incentive effect is only present in December and that at that time, travellers have rational expectations about the number of trips they make next year. For Silver, the expected number of flights is about 8 flights, whereas for Gold and Platinum, it is about 14 flights and 21 flights, respectively, see Table 2.[34]

Members just below the Platinum threshold expect to gain an additional benefit of Platinum of 8 - 22% of the ticket price over 10.5 return trip tickets in the next year (depending on whether tier level specific or pooled estimates are used in the calculation). Consequently, these members are willing to manipulate in order to get to a higher tier level (i.e., make additional flights) if the costs are less than about 84 – 231% of one return trip ticket. Assuming that the the additional return trip does not provide any utility to the members,[35] most members just below the Platinum threshold would be willing to manipulate by making one or two additional flights – as consistent with our findings of large (end-of-the-year) incentive effects at this threshold.

In contrast to Platinum travellers, we find much smaller incentive effects for Gold and particularly for Silver. This is consistent with that the expected benefits of a higher tier level in the next year are less than the costs of one

[33]Note however that our incentive effect estimates imply positive benefits for Gold, so the assumption that there are zero benefits of Gold implies irrational behavior.

[34]For Silver, for mileage and flight thresholds, the average number of flights are the same. For the other two thresholds, we use the mileage thresholds, as the consumption effect is based on these thresholds.

[35]This seems a reasonable assumption on average, given that for travellers who make a useful trip, this will be an overestimate of the cost, but for travellers who make this trip purely for incentive reasons this will be an underestimate due to time losses.

additional flight, i.e. the expected benefits are around 56 - 70% of the costs for members just below the Gold threshold and only 12 - 32% for members just below Silver.

7. Conclusion

In the current paper we quantify the switching costs induced by tier levels, which are an important characteristic of frequent flier programs and other loyalty programs. We derive these switching costs from an overlooked demand effect, which we label as the *consumption effect*: members with an elite tier level increase their demand with the airlines in the program to enjoy the tier level benefits.

We empirically estimate this consumption effect based on monthly flight data from a major international frequent flier program that uses discrete tier thresholds. A standard regression discontinuity design is however invalid, because of manipulating: members who are just below a tier threshold are likely to make additional flights to obtain a higher level. We therefore introduce a novel idea using a fuzzy regression discontinuity approach where we exploit that members only manipulate just before the end of the year. This idea can also be applied in other contexts where a regression discontinuity approach is intended, but where manipulation of the running variable is problematic for valid inference.

The estimated consumption effects imply switching costs equivalent to about 30 - 41% of the ticket price for the highest level, 13 - 17% for the middle level, and 3 - 8% for the lowest level. These findings are consistent with the notion that a convex reward structure is critical in the success of loyalty programs (Levine, 1987; Banerjee and Summers, 1987; Klemperer, 1995; Borenstein, 1996). This study demonstrates how airlines use tier levels to offer convex rewards to high-demand members, who are no longer affected by frequency rewards, and that the resulting switching costs are substantial.

References

Banerjee, A. V. and Summers, L. H. (1987). On frequent flyer programs and other loyalty-inducing economic arrangements. *Harvard Institute of Economic Research Discussion Paper*, 1337.

Basso, L. J., Clements, M. T., and Ross, T. W. (2009). Moral Hazard and Customer Loyalty Programs. *American Economic Journal: Microeconomics*, 1(1):101–123.

Basso, L. J., Feres, F., and Pezoa, R. A. (2016). Frequent Flyer Programs, Moral Hazard and Rewards per miles vs per dollar. *mimeo*.

Berry, S. and Jia, P. (2010). Tracing the woes: Tracing the woes: An empirical analysis of the airline industry. *American Economic Journal: Microeconomics*, 2(3):1–43.

Bertanha, M. (2020). Regression discontinuity design with many thresholds. *Journal of Econometrics*, 218(1):216–241.

Borenstein, S. (1996). Repeat Buyer Programs in Network industries. In *Networks, Infrastructure, and the New Task for Regulation*. University of Michigan Press, Ann Arbor: Michigan.

Calonico, S., Cattaneo, M. D., and Titiunik, R. (2014). Robust Nonparametric Confidence Intervals for Regression-Discontinuity Designs. *Econometrica*, 82(6):2295–2326.

Camacho, A. and Conover, E. (2011). Manipulation of Social Program Eligibility. *American Economic Journal: Economic Policy*, 3(2):41–65.

Cameron, A. C. and Trivedi, P. K. (2005). *Microeconometrics: methods and applications*. Cambridge University Press.

Cattaneo, M. D., Jansson, M., and Ma, X. (2018). Manipulation Testing Based on Density Discontinuity. *The Stata Journal: Promoting communications on statistics and Stata*, 18(1):234–261.

Cattaneo, M. D., Keele, L., Titiunik, R., and Vazquez-Bare, G. (2016). Interpreting regression discontinuity designs with multiple cutoffs. *Journal of Politics*, 78(4):1229–1248.

Chen, Y. and Ovchinnikov, A. (2019). Quantifying Mileage Runs. *SSRN Electronic Journal*.

Dong, Y. (2019). Regression Discontinuity Designs With Sample Selection. *Journal of Business & Economic Statistics*, 37(1):171–186.

Dowling, G. R. and Uncles, M. (1997). Do customer loyalty programs really work? *Sloan management review*, 38(4):71–82.

Frandsen, B. R. (2017). Party Bias in Union Representation Elections: Testing for Manipulation in the Regression Discontinuity Design when the Running Variable is Discrete. In *Regression Discontinuity Designs: Theory and Applications.*, pages 281–315. Emerald Publishing Limited.

Gans, J. S., Goldfarb, A., and Lederman, M. (2021). Exit, Tweets, and Loyalty. *American Economic Journal: Microeconomics*, 13(2):68–112.

Gerard, F., Rokkanen, M., and Rothe, C. (2020). Bounds on treatment effects in regression discontinuity designs with a manipulated running variable. *Quantitative Economics*, 11(3):839–870.

Hahn, J., Todd, P., and Klaauw, W. (2001). Identification and Estimation of Treatment Effects with a Regression-Discontinuity Design. *Econometrica*, 69(1):201–209.

Hartmann, W. R. and Viard, V. B. (2008). Do frequency reward programs create switching costs? A dynamic structural analysis of demand in a reward program. *Quantitative Marketing and Economics*, 6(2):109–137.

Imbens, G. and Wager, S. (2018). Optimized Regression Discontinuity Designs. *The Review of Economics and Statistics*, pages 1–15.

Imbens, G. W. and Lemieux, T. (2008). Regression discontinuity designs: A guide to practice. *Journal of Econometrics*, 142(2):615–635.

Jacob, B. A. and Lefgren, L. (2004). Remedial Education and Student Achievement: A Regression-Discontinuity Analysis. *Review of Economics and Statistics*, 86(1):226–244.

Klemperer, P. (1987). Markets with Consumer Switching Costs. *Quarterly Journal of Economics*, 102(2):375–394.

Klemperer, P. (1995). Competition when Consumers have Switching Costs: An Overview with Applications to Industrial Organization, Macroeconomics, and International Trade. *The Review of Economic Studies*, 62(4):515–539.

Kleven, H. J. and Waseem, M. (2013). Using notches to uncover optimization frictions and structural elasticities: Theory and evidence from Pakistan. *Quarterly Journal of Economics*, 128(2):669–723.

Kolesár, M. and Rothe, C. (2018). Inference in Regression Discontinuity Designs with a Discrete Running Variable. *American Economic Review*, 108(8):2277–2304.

Lederman, M. (2007). Do enhancements to loyalty programs affect demand? The impact of international frequent flyer partnerships on domestic airline demand. *The RAND Journal of Economics*, 38(4):1134–1158.

Lederman, M. (2008). Are Frequent-Flyer Programs a Cause of the "Hub Premium"? *Journal of Economics & Management Strategy*, 17(1):35–66.

Lee, D. S. and Lemieux, T. (2010). Regression Discontinuity Designs in Economics. *Journal of Economic Literature*, 48(2):281–355.

Levine, M. E. (1987). Airline Competition in Deregulated Markets: Theory, Firm Strategy, and Public Policy. *Yale Journal on Regulation*, 4:393–494.

Matsudaira, J. D. (2008). Mandatory summer school and student achievement. *Journal of Econometrics*, 142(2):829–850.

McCrary, J. (2008). Manipulation of the running variable in the regression discontinuity design: A density test. *Journal of Econometrics*, 142(2):698–714.

Nickell, S. (1981). Biases in dynamic panel models with fixed effects. *Econometrica*, 49(6):1417–1426.

Orhun, A. Y., Guo, T., and Hagemann, A. (2022). Reaching for Gold: Frequent-Flyer Status Incentives and Moral Hazard. *Marketing Science*.

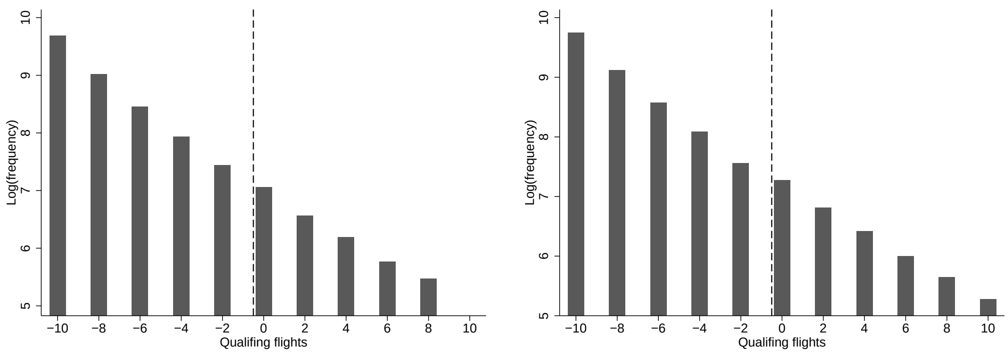
Sallee, J. M. and Slemrod, J. (2012). Car notches: Strategic automaker responses to fuel economy policy. *Journal of Public Economics*, 96(11-12):981–999.

Train, K. E. and Winston, C. (2007). Vehicle choice behavior and the declining market share of U.S. automakers. *International Economic Review*, 48(4):1469–1496.

Urquiola, M. and Verhoogen, E. (2009). Class-Size Caps, Sorting, and the Regression-Discontinuity Design. *American Economic Review*, 99(1):179–215.

Appendix A. Distributions of odd number of qualifying flights

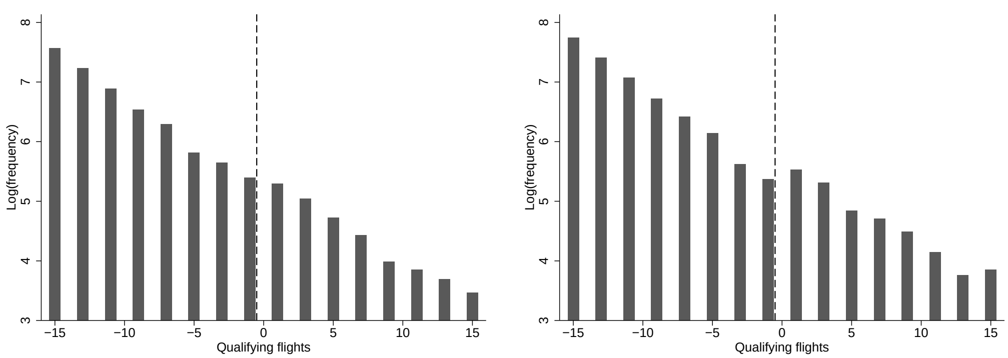
Figure A.1: Qualifying flights around Silver thresholds, odd numbers.



(a) November

(b) December

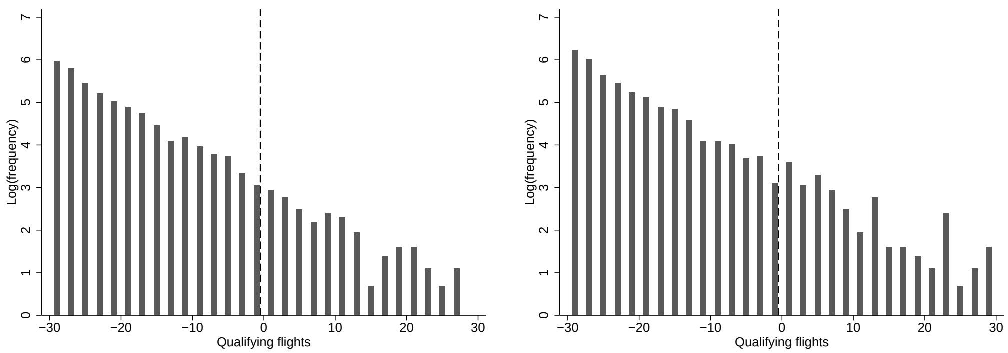
Figure A.2: Qualifying flights around Gold thresholds, odd numbers.



(a) November

(b) December

Figure A.3: Qualifying flights around Platinum thresholds, odd numbers.



(a) November (b) December

Appendix B. Incentive and manipulation tests

Table B.1: Manipulation tests for running variables miles (McCrary)

	Silver		Gold		Platinum	
	Nov	Dec	Nov	Dec	Nov	Dec
McCrary (estimate)	-0.022	0.021	0.100	0.157	0.146	0.324
McCrary (se)	0.056	0.051	0.074	0.068	0.097	0.091
McCrary (p-value)	0.6908	0.6825	0.1761	0.0205	0.1320	0.0004
Bandwidth	15,000	15,000	15,000	15,000	30,000	30,000

Table B.2: Manipulation tests for running variables flights (Frandsen, k= 0)

	Silver		Gold		Platinum	
	Nov	Dec	Nov	Dec	Nov	Dec
Frandsen (p-value, even)	.0916	4.6e-10	.2636	1.9e-09	.9794	9.2e-07
Frandsen (p-value, odd)	0.2298	0.4978	0.4868	0.0164	0.8019	0.0174
Bonferroni (p-value)	0.18	0.00	0.52	0.00	1.00	0.00

Appendix C. Tests for discrete running variables

Appendix C.1. Calculation of k for the Frandsen test

The Frandsen test requires one to set k. To determine this value, we assume that number of flights is a discretized version of an exponential distribution, where $f(x) = \lambda e^{-\lambda x}$, with parameter λ. The degree of curvature, as a function of Δ, corresponds to a k, defined as Frandsen (2017):

$$k = \frac{f(x+\Delta) + f(x-\Delta) - 2f(x)}{f(x+\Delta) + f(x-\Delta)},$$

where Δ is the length of the interval between points. Given the exponential function, it can be written as:

$$k = \frac{e^{\lambda\Delta} + e^{-\lambda\Delta} - 2}{e^{\lambda\Delta} + e^{-\lambda\Delta}}.$$

In our application, x refers to the number of flights, and we make a distinction between odd and even frequencies, so $\Delta = 2$. In this case:

$$k = 1 - \frac{2}{e^{\lambda\Delta} + e^{-\lambda\Delta}} \approx 2\lambda^2,$$

where the latter approximation holds when λ is close to zero.

We estimate λ using data about the distribution of the number of flights, while excluding data points at and around the threshold of interest (i.e. one support point at the threshold and two support points immediately adjacent to the threshold), see Table C.1. It appears that for Silver, λ is around

Table C.1: Alternative Frandsen test running variable flights

	p (k = 0)	λ	k	p (k > 0)
Silver Nov, odd	0.230	0.231	0.107	0.975
Silver Nov, even	0.092	0.249	0.124	0.972
Silver Dec, odd	0.498	0.219	0.096	0.986
Silver Dec, even	0.000	0.242	0.117	0.006
Gold Nov, odd	0.487	0.136	0.037	0.571
Gold Nov, even	0.264	0.145	0.042	0.345
Gold Dec, odd	0.016	0.131	0.034	0.031
Gold Dec, even	0.000	0.137	0.037	0.000
Platinum Nov, odd	0.802	0.092	0.017	1.000
Platinum Nov, even	0.979	0.093	0.017	1.000
Platinum Dec, odd	0.017	0.088	0.016	0.032
Platinum Dec, even	0.000	0.090	0.016	0.000

0.22-0.25, for Gold, it is 0.13-0.14 and for Platinum around 0.09. Given these estimates, we calculate k. Given k we calculate the p-values, which are reported in the last column of the Table C.1. For comparison, we report again the values for $k = 0$, as reported in Appendix B. We find that for the $k > 0$, p-values are sometimes substantially larger, but for none of the tests, the p-value increases from below 0.05 to values above 0.05

Appendix C.2. Alternative test

The Frandsen test starts from a smoothness assumption which implies linearity and then requires one to specify k to allow for deviations of linearity. We introduce here an alternative approach which assumes that number of flights, x, is a *local* discretized version of an exponential distribution, where $f(x) = \lambda e^{-\lambda x}$, with arbitrary and unknown parameter λ. We focus on *three* support points of the running variable distribution. One support point at the threshold and two support points immediately adjacent to the threshold. We observe the frequency $\hat{f}(.)$ of three support points: $t - \Delta$, t, $t + \Delta$, where t refers to the threshold value (e.g., 15 flights) and where Δ denotes the length of the interval between points (in our application, two flights).

Let us now introduce Q, defined as the *log-weighted* frequency *share* at the threshold, defined as follows:

$$Q = \frac{\log(\hat{f}(t))}{\log(\hat{f}(t-\Delta)) + \log(\hat{f}(t)) + \log(\hat{f}(t-\Delta))}.$$

It is straightforward to show that given the exponential distribution assumption, the null hypothesis of no manipulation implies that Q = 1/3. The result is intuitive, as the exponential distribution implies that the logarithm of the frequency is a linear function of x, hence one third of the log of observations should be at the threshold. Hence, we will test for the null hypothesis that Q = 1/3. The statistical properties of Q are unknown, consequently p-values are obtained by bootstrapping, as reported in the last column of the Table C.2.

For Gold and Platinum these p-values yield the same conclusions as the (alternative) Frandsen test: there is manipulation in December, but not in November. The only exception is the alternative test results for flights around the Silver threshold. However, the amount of manipulating as indicated by the test is negligible – i.e. at the threshold, the frequency is only slightly

Table C.2: Alternative test running variable tests

	Q	Q - 1/3	s.e.	t-value	p-value
Silver, Nov, odd	0.338	0.004	0.001	8.294	0.000
Silver, Nov, even	0.335	0.001	0.001	2.478	0.007
Silver, Dec, odd	0.341	0.008	0.001	15.450	0.000
Silver, Dec, even	0.336	0.003	0.001	5.090	0.000
Gold, Nov, odd	0.331	-0.003	0.001	-1.816	0.965
Gold, Nov, even	0.336	0.003	0.002	1.463	0.072
Gold, Dec, odd	0.347	0.013	0.001	10.948	0.000
Gold, Dec, even	0.340	0.007	0.002	4.598	0.000
Platinum, Nov, odd	0.332	-0.001	0.007	-0.199	0.579
Platinum, Nov, even	0.335	0.001	0.012	0.110	0.456
Platinum, Dec, odd	0.365	0.031	0.005	6.390	0.000
Platinum, Dec, even	0.353	0.019	0.006	3.122	0.001

larger than it's theoretical value given no manipulation. Our finding of a rejection of the null hypothesis of no manipulation combined with a small size effect makes sense, because we have nearly 8,000 observations near this threshold, so even a small deviation from the exponential distribution assumption is likely to induce one to falsely reject the null hypothesis of no manipulation.

Appendix D. Consumption analysis: first stage estimates

Table D.1: First stage estimation results for running variable mileage

	Silver Dec	Gold Nov	Platinum Nov
Threshold	0.733***	0.499***	0.403***
	(0.0132)	(0.0264)	(0.0255)
Mileage	0.340***	0.586***	0.476***
	(0.0685)	(0.147)	(0.0407)
Mileage^2	0.238***	0.294*	0.109***
	(0.0802)	(0.179)	(0.0149)
Treatment * Mileage	-0.341***	-0.587***	-0.477***
	(0.0685)	(0.147)	(0.0407)
Treatment * Mileage^2	-0.237***	-0.294*	-0.108***
	(0.0802)	(0.179)	(0.0149)
Year dummy (2014)	0.00542	0.0136	-0.0104
	(0.00461)	(0.00925)	(0.00748)
Constant	0.264***	0.494***	0.603***
	(0.0133)	(0.0268)	(0.0257)
Nr. of observations	17,316	5,237	5,698
Adjusted R^2	0.589	0.497	0.625

Table D.2: First stage estimation results for running variable flights

	Silver Nov
Threshold	0.387***
	(0.00996)
Flights	0.116***
	(0.00296)
Flights^2	0.00591***
	(0.000210)
Treatment * Flights	-0.116***
	(0.00296)
Treatment * Flights^2	-0.00591***
	(0.000210)
Year dummy (2014)	-0.00691***
	(0.00180)
Constant	0.617***
	(0.01000)
Nr. of observations	87,483
Adjusted R^2	0.473

Appendix E. Consumption effect analysis: full estimates

Table E.3: Consumption effect on flights of qualifying on miles

	Silver Dec		Gold Nov		Platinum Nov	
	2SLS	IVPoisson	2SLS	IVPoisson	2SLS	IVPoisson
Treatment	0.460	0.0760	2.484	0.358	7.920**	0.534**
	(0.646)	(0.147)	(1.849)	(0.245)	(3.858)	(0.219)
Mileage	6.351***	0.983***	-3.065	-0.304	-2.018	-0.156
	(1.680)	(0.269)	(4.259)	(0.398)	(2.743)	(0.159)
Mileage^2	3.854**	0.529*	-4.220	-0.313	-0.605	-0.0400
	(1.887)	(0.304)	(4.714)	(0.395)	(0.815)	(0.0472)
Treatment * Mileage	-7.469**	-1.114***	4.503	0.436	-1.783	-0.0241
	(2.906)	(0.393)	(5.800)	(0.495)	(2.812)	(0.155)
Treatment * Mileage^2	-1.132	-0.202	4.190	0.292	2.542	0.132
	(4.010)	(0.524)	(8.119)	(0.634)	(1.794)	(0.0890)
Year dummy (2014)	0.0791	0.00549	0.429	0.0333	-0.411	-0.0177
	(0.124)	(0.0204)	(0.317)	(0.0281)	(0.388)	(0.0219)
Constant	7.335***	1.979***	9.290***	2.107***	13.63***	2.535***
	(0.429)	(0.117)	(1.399)	(0.204)	(2.790)	(0.171)
Nr. of observations	17,316		5,237		5,698	

Table E.4: Consumption effect on miles of qualifying on miles

	Silver Dec		Gold Nov		Platinum Nov	
	2SLS	IVPoisson	2SLS	IVPoisson	2SLS	IVPoisson
Treatment	246.4	0.0330	468.8	0.00129	4752.0	0.192
	(900.1)	(0.113)	(2826.9)	(0.246)	(5132.8)	(0.214)
Mileage	6653.1***	0.633***	9482.3	0.558	2711.1	0.0693
	(2328.8)	(0.239)	(6625.4)	(0.395)	(4096.7)	(0.156)
Mileage^2	1893.3	0.0429	7496.9	0.436	229.2	0.00242
	(2606.2)	(0.276)	(7269.1)	(0.400)	(1241.4)	(0.0440)
Treat. * Mileage	-371.3	-0.0949	-10435.3	-0.541	-1125.5	-0.0141
	(4241.2)	(0.371)	(9028.0)	(0.472)	(4375.6)	(0.147)
Treat. * Mileage^2	-3784.9	-0.258	1715.9	-0.0944	529.0	0.0115
	(6000.4)	(0.480)	(12822.1)	(0.614)	(2434.3)	(0.0722)
Year dum. (2014)	62.88	-0.000059	69.08	0.00335	-909.3	-0.0288
	(181.7)	(0.0194)	(476.4)	(0.0259)	(572.2)	(0.0193)
Constant	10769.1***	9.279***	19319.1***	9.886***	30530.8***	10.28***
	(596.7)	(0.0858)	(2191.4)	(0.212)	(3982.8)	(0.184)
Nr. of observations	17,316		5,237		5,698	

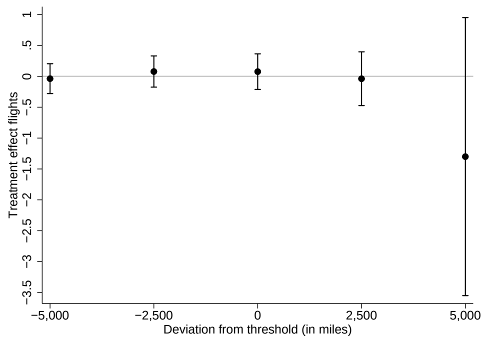
Table E.5: Consumption effect on flights of qualifying on flights

	Silver Nov	
	2SLS	IVPoisson
Treatment	0.0171	0.0753
	(0.782)	(0.130)
Flights	1.096***	0.0766***
	(0.132)	(0.0231)
Flights^2	0.0273***	-0.00412***
	(0.00772)	(0.00133)
Treatment * Flights	-0.203	-0.00689
	(0.167)	(0.0216)
Treatment * Flights^2	-0.0422*	0.00205
	(0.0232)	(0.00206)
Year dummy (2014)	-0.139***	-0.0297***
	(0.0440)	(0.00798)
Constant	11.65***	2.395***
	(0.617)	(0.119)
Observations	87,483	

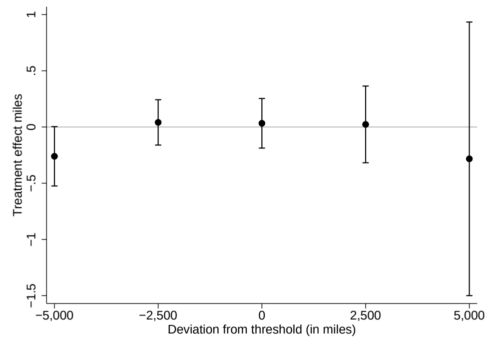
Table E.6: Consumption effect on miles of qualifying on flights

	Silver Nov	
	2SLS	IVPoisson
Treatment	-856.1	-0.0294
	(1004.8)	(0.252)
Flights	969.0***	0.0639*
	(174.3)	(0.0352)
Flights^2	16.48	-0.00488***
	(10.23)	(0.00171)
Treatment * Flights	-283.2	-0.00569
	(207.2)	(0.0324)
Treatment * Flights^2	-43.87	0.00222
	(27.06)	(0.00271)
Year dummy (2014)	-292.5***	-0.0544***
	(59.64)	(0.0108)
Constant	11789.8***	9.346***
	(805.9)	(0.237)
Observations	87,483	

Appendix F. Placebo plots

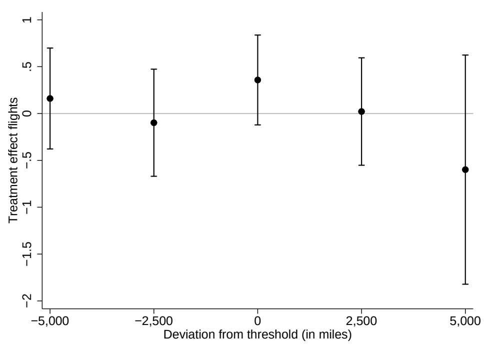


(a) Flights

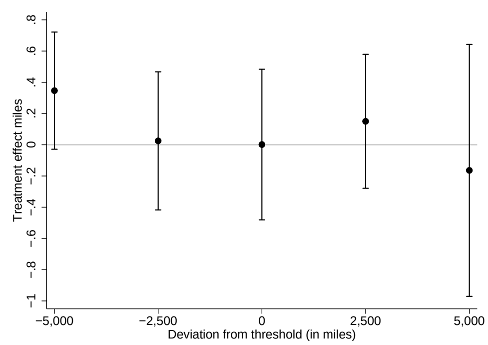


(b) Miles

Figure F.1: Silver (December), running variables miles (Poisson)

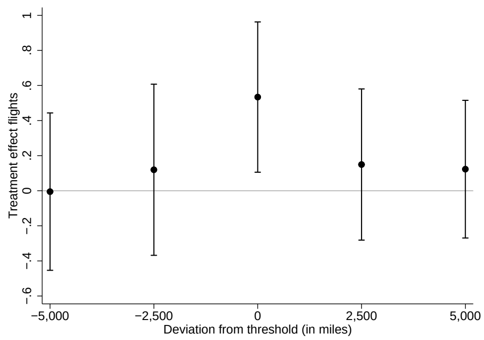


(a) Flights

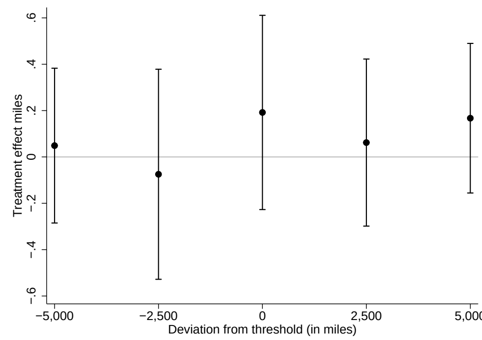


(b) Miles

Figure F.2: Gold (November), running variables miles (Poisson)

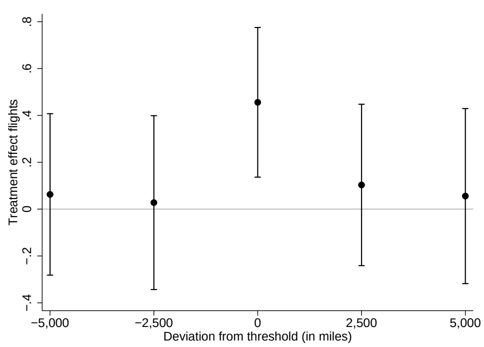


(a) Flights

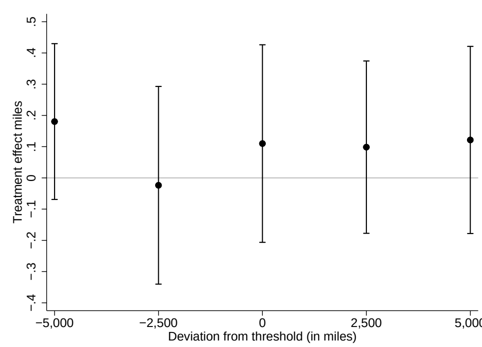


(b) Miles

Figure F.3: Platinum (November), running variables miles (Poisson)

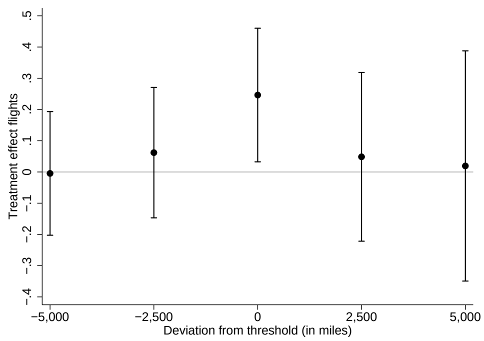


(a) Flights

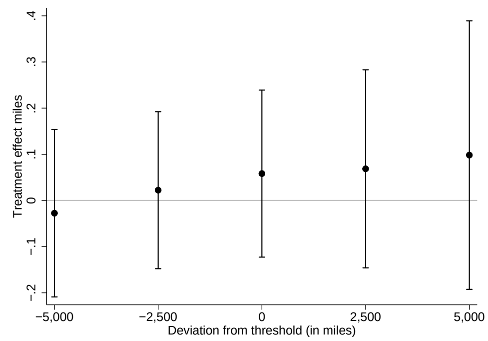


(b) Miles

Figure F.4: Gold and Platinum (November), running variables miles (Poisson)

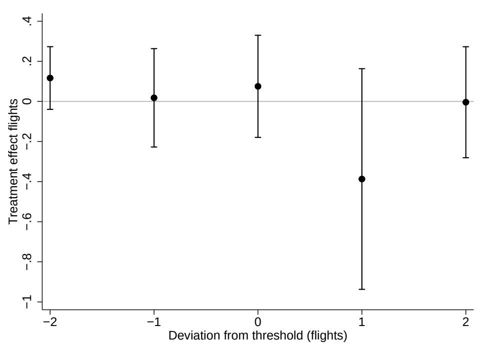


(a) Flights

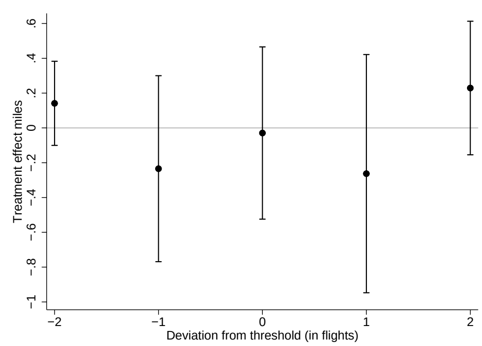


(b) Miles

Figure F.5: Silver (December), Gold and Platinum (November), running variables miles (Poisson)



(a) Flights



(b) Miles

Figure F.6: Silver (November), running variables flights (Poisson)

Appendix G. Intensive and extensive margins

Appendix G.1. Extensive margin

Table G.1: Extensive margin, qualifying via miles

	Silver Dec	Gold Nov	Platinum Nov
Treatment	0.0473	0.126**	0.0355
	(0.0306)	(0.0589)	(0.0498)
Mileage	0.264***	0.00478	0.0162
	(0.0898)	(0.147)	(0.0461)
Mileage^2	0.144	0.101	0.00430
	(0.106)	(0.163)	(0.0148)
Treatment * Mileage	-0.348**	-0.297	-0.0260
	(0.138)	(0.184)	(0.0474)
Treatment * Mileage^2	0.0258	0.340	0.00278
	(0.188)	(0.244)	(0.0223)
Year dummy (2014)	-0.000556	0.0175*	-0.00480
	(0.00698)	(0.00980)	(0.00673)
Constant	0.724***	0.763***	0.928***
	(0.0215)	(0.0471)	(0.0419)
Nr. of observations	17,316	5,237	5,698
Adjusted R^2	0.031	0.044	0.014

Table G.2: Extensive margin, qualifying via flights

	Silver Nov
Treatment	-0.00282
	(0.0275)
Flights	0.00699
	(0.00564)
Flights^2	-0.00230***
	(0.000365)
Treatment * Flights	0.00419
	(0.00589)
Treatment * Flights^2	0.00145**
	(0.000620)
Year dummy (2014)	-0.0128***
	(0.00298)
Constant	0.903***
	(0.0237)
Nr. of observations	87,483
Adjusted R^2	0.048

Appendix G.2. Intensive margin

Table G.3: Consumption effect on flights of qualifying on miles intensive margin, linear

	Silver Dec 2SLS	Gold Nov 2SLS	Platinum Nov 2SLS
Treatment	-0.107	1.064	8.128*
	(0.862)	(2.357)	(4.271)
Mileage	5.636***	-3.780	-2.783
	(2.185)	(4.956)	(3.024)
Mileage^2	3.305	-6.515	-0.800
	(2.477)	(5.231)	(0.876)
Treatment * Mileage	-5.952*	10.26	-0.975
	(3.469)	(6.310)	(2.888)
Treatment * Mileage^2	-2.064	-0.785	2.664
	(4.754)	(8.633)	(1.871)
Year dummy (2014)	0.109	0.223	-0.343
	(0.160)	(0.348)	(0.399)
Constant	10.21***	12.15***	14.26***
	(0.612)	(1.883)	(3.177)
Nr. of observations	11,783	4,427	5,304

Table G.4: Consumption effect on flights of qualifying on miles intensive margin, log

	Silver Dec 2SLS	Gold Nov 2SLS	Platinum Nov 2SLS
Treatment	0.00298	0.0591	0.172
	(0.0802)	(0.184)	(0.202)
Mileage	0.701***	-0.00756	0.0587
	(0.207)	(0.392)	(0.163)
Mileage^2	0.497**	-0.188	0.0107
	(0.240)	(0.416)	(0.0497)
Treatment * Mileage	-0.666**	0.179	-0.143
	(0.325)	(0.489)	(0.163)
Treatment * Mileage^2	-0.482	0.0610	0.0453
	(0.445)	(0.663)	(0.0876)
Year dummy (2014)	0.0216	0.0129	-0.00967
	(0.0158)	(0.0264)	(0.0227)
Constant	1.956***	2.213***	2.592***
	(0.0567)	(0.148)	(0.161)
Nr. of observations	11,783	4,427	5,304

Table G.5: Consumption effect on flights of qualifying on miles intensive margin, Poisson

	Silver Dec 2SLS	Gold Nov 2SLS	Platinum Nov 2SLS
Treatment	-0.0261	0.123	0.493**
	(0.138)	(0.253)	(0.216)
Mileage	0.603**	-0.288	-0.173
	(0.244)	(0.373)	(0.153)
Mileage^2	0.322	-0.476	-0.0450
	(0.273)	(0.366)	(0.0449)
Treatment * Mileage	-0.631*	0.743	0.00287
	(0.351)	(0.458)	(0.147)
Treatment * Mileage^2	-0.206	-0.0339	0.130
	(0.467)	(0.578)	(0.0863)
Year dummy (2014)	0.0101	0.0172	-0.0132
	(0.0179)	(0.0267)	(0.0207)
Constant	2.338***	2.459***	2.612***
	(0.111)	(0.215)	(0.169)
Nr. of observations	11,783	4,427	5,304

Table G.6: Consumption effect on miles of qualifying on miles intensive margin, linear

	Silver Dec 2SLS	Gold Nov 2SLS	Platinum Nov 2SLS
Treatment	-719.8	-3546.8	3720.3
	(1173.4)	(3535.0)	(5556.4)
Mileage	4109.8	12200.1	2416.3
	(2936.8)	(7617.7)	(4392.7)
Mileage^2	-964.5	6906.9	108.3
	(3312.1)	(8008.1)	(1302.1)
Treatment * Mileage	5939.8	-5000.0	-311.4
	(4930.4)	(9671.2)	(4465.7)
Treatment * Mileage^2	-5339.4	-8972.2	359.9
	(6943.1)	(13417.8)	(2479.8)
Year dummy (2014)	71.41	-377.6	-828.4
	(229.0)	(517.0)	(576.3)
Constant	14992.9***	25758.3***	32897.6***
	(828.8)	(2888.1)	(4409.5)
Nr. of observations	11,783	4,427	5,304

Table G.7: Consumption effect on miles of qualifying on miles intensive margin, log

	Silver Dec 2SLS	Gold Nov 2SLS	Platinum Nov 2SLS
Treatment	-0.0504	-0.239	0.193
	(0.102)	(0.229)	(0.217)
Mileage	0.575**	0.832*	0.0199
	(0.271)	(0.494)	(0.182)
Mileage^2	0.170	0.509	-0.0196
	(0.316)	(0.521)	(0.0559)
Treatment * Mileage	0.00655	-0.737	-0.0323
	(0.416)	(0.605)	(0.184)
Treatment * Mileage^2	-0.545	-0.319	0.0621
	(0.568)	(0.810)	(0.0945)
Year dummy (2014)	0.00867	-0.0117	0.0000938
	(0.0210)	(0.0327)	(0.0251)
Constant	9.199***	9.887***	10.04***
	(0.0729)	(0.187)	(0.177)
Nr. of observations	11,783	4,427	5,304

Table G.8: Consumption effect on miles of qualifying on miles intensive margin, Poisson

	Silver Dec 2SLS	Gold Nov 2SLS	Platinum Nov 2SLS
Treatment	-0.0528	-0.254	0.141
	(0.101)	(0.254)	(0.209)
Mileage	0.247	0.586	0.0579
	(0.207)	(0.374)	(0.150)
Mileage^2	-0.167	0.297	-0.00150
	(0.236)	(0.376)	(0.0415)
Treatment * Mileage	0.396	-0.250	0.00673
	(0.322)	(0.437)	(0.140)
Treatment * Mileage^2	-0.263	-0.433	0.00847
	(0.417)	(0.561)	(0.0686)
Year dummy (2014)	0.00360	-0.0153	-0.0243
	(0.0165)	(0.0244)	(0.0180)
Constant	9.623***	10.26***	10.37***
	(0.0777)	(0.224)	(0.180)
Nr. of observations	11,783	4,427	5,304