Evidence from Implementing a Novel Ticketing Solution

Report

Silvio Sticher
Hannes Wallimann

January 2026

Project funded by SBB research fund

Acknowledgment: We are grateful to the SBB Research Fund for financial support. We have benefited from comments by Kevin Blättler, Julian Frank, Nathalie Frischknecht, and Widar von Arx.

University of Applied Sciences and Arts Lucerne, Competence Center for Mobility

Contents

List of Figures **2**

List of Tables **2**

Management Summary (deutsch) **3**

1 Introduction **5**

2 Literature Review **6**
2.1 Implementation of Public-Transportation Pricing Policies 6
2.2 Mechanisms Affecting Generalizability of Pilot-Study Findings 7

3 Background **8**

4 Research Design **11**
4.1 Overall Research Design 11
4.2 Framework for Causal Effect Estimation (Steps 1 and 2) 12
4.3 Framework for Bias Identification and Correction (Steps 3 and 4) 13

5 Data **14**
5.1 Final Product Data 14
5.2 Pilot Study Data 15
5.3 Survey Data 15

6 Results **17**
6.1 Causal Effect of the Final Product 17
6.2 Initial Causal Effect of the Pilot Study 19
6.3 Identifying Selection Bias in the Pilot Study 20
6.4 De-Biased Causal Effect of the Pilot Study 21

7 Discussion and Conclusion **24**

A Causal Effect of the Final Product, Robustness Check **29**

B Additional Tables **31**

C Additional Figures **33**

List of Figures

1 Overall research design . . . 11
2 Final Product: yearly expenditure composition vs. control group . . . 18
3 Final Product (early adopters): yearly expenditure composition vs. control group 19
4 Pilot Study vs. Final Product: survey responses . . . 20
5 Propensity score distributions for the pilot study and final product groups (left panel: flexible model; right panel: parsimonious model). Scores represent the estimated probability of belonging to the final product group. . . . 23
6 De-biasing reduces the estimated selection bias (difference between Pilot Study and Final Product) by 51% (parsimonious model). . . . 23
A.1 Cumulative number of Final Product customers (sample observations) until December 2024 . . . 29
C.1 Pilot Study vs. Final Product: age distribution . . . 33
C.2 Pilot Study vs. Final Product: variable importance (Random Forest classification) 33

List of Tables

1 Potential mechanisms of a pilot-study bias . . . 8
2 Half Fare Travelcard PLUS (Final Product) and Pilot Study package options . . . 10
3 Stated vs. revealed participation . . . 16
4 Final Product: yearly incremental expenditure . . . 17
5 Pilot Study: yearly incremental expenditure (initial estimate) . . . 20
6 Flexible model: analysis of deviance (type II Wald χ^2 tests) . . . 22
8 De-biased yearly incremental expenditure (Pilot Study group) . . . 22
7 Parsimonious model: regression output (binary logit) . . . 23
A.1 Final Product: weekly incremental expenditure (robustness check) . . . 30
B.1 Covariate balance before and after matching (Final Product, pooled across packages) . . . 31
B.2 Pilot Study vs. Final Product: Mechanisms, survey statements, and hypotheses . 32

Management Summary (deutsch)

Der Schweizer Öffentliche Verkehr (ÖV) entwickelt neue und innovative Tarifmodelle, um Kundenbedürfnissen gerecht zu werden und gleichzeitig die Einnahmen zu sichern. Um die Auswirkungen solcher Angebote vor einer Markteinführung abzuschätzen, verlässt sich die Branche häufig auf Pilotstudien (Markttests). Es stellt sich jedoch die zentrale Frage, inwieweit die Ergebnisse solcher Pilotstudien auf den Gesamtmarkt übertragbar sind.

In dieser Studie untersuchen wir diese Frage am Beispiel des «Halbtax PLUS», einem im Dezember 2023 eingeführten Guthaben-Abonnement. Wir vergleichen die Ergebnisse des vorgelagerten Markttests (2021–2023) mit den tatsächlichen Auswirkungen nach der Markteinführung. Unsere Analyse basiert auf ÖV-Daten (Abos und Einzelbillette) von über 100'000 Kundinnen und Kunden sowie einer ergänzenden Befragung von rund 1'000 Teilnehmenden, welche am Markttest teilgenommen haben oder das Halbtax PLUS kauften (oder beides).

Zentrale Ergebnisse:

- **Diskrepanz bei den Mehreinnahmen:** Die Teilnehmenden des Markttests generierten durchschnittliche Mehreinnahmen von MU 77.47[1] pro Person (Replikation der Studie von Sticher and Blättler, 2024, mit neuer Methodik). Nach der definitiven Markteinführung fiel dieser Effekt jedoch deutlich geringer aus: Über alle Kundensegmente hinweg betrug der durchschnittliche Mehrumsatz lediglich MU 15.75 pro Person und Jahr.
- **Selbstselektion als Hauptursache:** Unsere Analyse zeigt, dass sich die Teilnehmenden des Markttests systematisch von den späteren Kundinnen und Kunden unterscheiden: Sie zeichnen sich durch eine signifikant höhere Innovationsfreude und eine höhere Risikoaffinität aus. Diese Eigenschaften korrelieren im vorliegenden Fall mit zusätzlichen Ausgaben, womit die Mehreinnahmen des Markttests nicht vollumfänglich auf die Markteinführung übertragbar sind.
- **Korrektur durch Gewichtung möglich:** In einem "Proof of Concept" zeigen wir, dass sich diese Verzerrung korrigieren lässt. Durch eine statistische Gewichtung, die auf nur zwei einfachen Fragen zu Innovationsfreude und Risikobereitschaft basiert, können wir die geschätzten Mehreinnahmen im Markttest auf MU 39.59 nach unten korrigieren.
- **Verbleibender Produkteffekt:** Auch nach der Korrektur verbleibt eine zu optimistische Prognose für die Mehreinnahmen. Wir führen dies auf Anpassungen am Produktdesign zurück, so etwa auf neue Paketgrössen und Preise für Jugendliche, die Ausweitung vom Automatischen Ticketing zu einem breiteren Online/Mobile-Sortiment und den Beibezug von Sparbilletten.

Fazit für die Praxis: Die Ergebnisse dieser Studie zeigen, dass Transportunternehmen und andere Entscheidungsträger die Resultate von Pilotstudien im ÖV nicht ungefiltert auf den Gesamtmarkt hochrechnen sollten. Die Teilnehmenden solcher Tests sind keine repräsentative Auswahl der Gesamtbevölkerung, sondern eine spezifische Gruppe von Personen.

Dies bedeutet jedoch nicht, dass Pilotstudien wertlos sind – im Gegenteil. Unsere Arbeit zeigt einen praktikablen Weg auf, wie diese Verzerrungen korrigiert werden können. Durch die

[1] Aus Gründen der geschäftlichen Vertraulichkeit haben wir in vorliegender Studie sämtliche Finanzbeträge in anonymisierte Währungseinheiten (MU, für "Monetary Units") umgerechnet. Wir haben dazu eine skalare Transformation angewendet; folglich bleiben relative Differenzen, prozentuale Änderungen und die statistische Signifikanz von dieser Transformation unberührt.

gezielte Erhebung weniger zusätzlicher Merkmale (Innovationsfreude, Risikoaffinität) haben Unternehmen die Möglichkeit, ihre Prognosen zu verbessern. Dies ist entscheidend, um realistische Erwartungen an neue Tarifprodukte zu stellen.

1 Introduction

The effects of public-transportation pricing policies are highly relevant for both public transportation agencies and policymakers, not least because of their profound impact on both passenger demand and agency revenues. Consequently, a broad research literature has emerged on the ex post impact evaluation of such policies (see, e.g., Wallimann et al., 2023; Schlett and Loder, 2024; Waldorf et al., 2025). On the other hand, the effects of pricing policies can be estimated ex ante through surveys or pilot studies (see, e.g., Sticher and Blättler, 2024; Weibel et al., 2024; Wallimann and Sticher, 2025). However, an important question remains: do pilot-study participants represent a distinct group whose behavior differs from that of buyers after the full market launch? This raises, to the best of our knowledge, an as yet unanswered question about the generalizability of pilot study findings for new public-transportation pricing policies.

To fill this research gap, we empirically analyze the Half Fare Travelcard PLUS, introduced by Swiss public-transportation providers in December 2023. We use this product as a case study to examine the generalizability of pilot-study findings. This product offers an allowance to purchase single tickets over the course of one year for a fixed price (below the allowance amount).

First, we estimate the effect of the final product's market launch on customer expenditure. We apply matching methods to compare 50,000 randomly selected buyers with a control group of 50,000 individuals not holding a Half Fare Travelcard PLUS. We match these groups on pre-treatment consumption patterns and socio-demographic characteristics, while also differencing out pre-treatment expenditure to isolate the incremental effect. Our primary analysis, which accounts for heterogeneous effects across different allowance packages, yields an incremental expenditure of MU 15.75[2] per customer and year. This result represents a significant reduction compared to the pilot-study evaluation of Sticher and Blättler (2024). When we pool customer segments match the basis of the pilot study, incremental expenditure drops even further to MU 2.90, and the gap becomes even more pronounced.

Second, we focus on the original pilot-study customers whom we could also reach for a dedicated survey. We replicate the finding of Sticher and Blättler (2024), estimating the incremental expenditure in the pilot study at MU 77.47. This point estimate is broadly consistent with the original evaluation despite the fact that we have to employ a different econometric approach due to differences in the setting (e.g., tariff periods of customers not running in sync).

Our central hypothesis is that—while product differences matter (the product was further developed after the pilot-study phase)—self-selection into the pilot-study group is a main driver of this stark divergence in incremental expenditure.

In a third step, we therefore survey both pilot study participants and buyers of the final product to identify potential mechanisms that may cause pilot-study participants to be a non-representative group—such as innovator traits (see, e.g., Rogers, 2003; Parasuraman and Colby, 2015), late-adopter traits (see, e.g., Vannoy and Palvia, 2010; Yuen et al., 2023), risk aversion (see, e.g., Santana and Parigi, 2015; Yoo and Chavas, 2021), and price sensitivity (see, e.g., Goldsmith et al., 2005). Based on the survey results, we identify which factors help explain differences between the two groups. We show that participants who describe themselves as being among the first to adopt new technologies or products, or as having high risk affinity, are more likely to be in the pilot-study group than the final-product group.

Finally, as a proof of concept, we re-estimate the effect of the pilot study. We apply propensity-odds weighting to the pilot-study group, adjusting their self-selection characteristics

[2] For reasons of commercial confidentiality, all financial figures in this study have been converted from Swiss Francs to anonymized Monetary Units (MU). We applied a constant scalar transformation to all expenditure data; consequently, all relative differences, percentage changes, and levels of statistical significance remain unaffected by this transformation.

to match the distribution of the final product's customers. Doing so, our estimates of incremental expenditure effects are reduced to MU 33.98 or MU 39.10, depending on whether we apply a flexible or a parsimonious model, respectively. In doing so, we both illustrate a method for addressing potential biases in pilot studies and provide practical guidance for public-transportation enterprises and policymakers on which variables are relevant for weighting pilot-study participants.

We proceed as follows. In Section 2, we review the research literature on public-transportation pricing policies and provide an overview of potential mechanisms that cause pilot-study participants to differ from buyers of the final product after its market launch. In Section 3, we outline the public-transportation tariff landscape in Switzerland and describe the Half Fare Travelcard Plus and its pilot study. In Section 4, we present our research design. We then describe our multiple data sets in Section 5. In Section 6, we report the findings from our analytical steps. Finally, in Section 7, we discuss our findings and offer our conclusions.

2 Literature Review

In this section, we review the two main bodies of literature that motivate our research. First, we discuss the research on the implementation and ex post evaluation of public-transportation pricing policies. Second, we establish the theoretical foundation for our hypothesis by reviewing the literature on potential mechanisms that may cause a pilot-study group to be non-representative.

2.1 Implementation of Public-Transportation Pricing Policies

Public-transportation passengers value many supply attributes, including service frequency, reliability, and affordable fares (Redman et al., 2013). Among these, fares are often the most direct lever for public-transportation agencies and operators to influence, as price changes do not require new infrastructure construction.

On the one hand, changes to public-transportation pricing policies may be driven by political decisions. Recent prominent examples include nationwide flat-rate tickets, such as Germany's 9-Euro-Ticket and its successor, the Deutschlandticket (Loder et al., 2024; Waldorf et al., 2025; Schlett and Loder, 2024), or Austria's KlimaTicket (Wallimann, 2024). Other political approaches range from the complete elimination of fares, as seen in Tallinn (Kębłowski, 2020; Cats et al., 2014), to price reductions driven by popular referendums, such as in Geneva (Wallimann et al., 2023).

On the other hand, public-transportation enterprises and agencies often trigger pricing-policy changes on their own initiative. These changes can range from fundamental shifts in tariff structures to the introduction of new products. An example of the former is Stockholm's 2017 move from a zonal to a flat-fare basis, which aimed to simplify the tariff system (Kholodov et al., 2021). The Half Fare Travelcard PLUS—a public-transportation credit introduced in Switzerland in 2023—is an example of the latter. This product was designed to elicit additional ridership and revenue from high-mobility customers who fall "below" the subscription threshold (see Sticher and Blättler (2024) and Weibel et al. (2024)). (We discuss the Half Fare Travelcard PLUS in greater detail in Section 3.)

Often, and not least for customer-relationship reasons, public-transportation pricing policies include discounts and additional offers. As public-transportation demand is generally inelastic at current prices (see, e.g., Holmgren, 2007), these discounts must be precisely-targeted and fenced-off, or they increase the need for subsidies. Therefore, from both an industry viewpoint and a welfare-economics perspective (Parry and Small, 2009), it is important to know how public-transportation policies affect key performance indicators, i.e., demand and revenue. For

data-driven decision-making, two possibilities exist: Policymakers can either rely on ex post impact evaluations of other, comparable pricing policies. However, this is fraught with risk. External validity is questionable if the underlying circumstances—such as the pricing-policy details, the market structure, or customer behavior—are not comparable to the original case (see, e.g., discussion by Wallimann et al., 2023). Alternatively, the public-transportation industry (together with public authorities) can estimate the impacts of policies ex ante. They can do this through surveys (see, e.g., Weibel et al., 2024) or through pilot studies (Sticher and Blättler, 2024); the latter offers the advantage of observing revealed preferences instead of only stated preferences.

Despite these advantages, pilot studies come with their own caveats. In particular, it is plausible that pilot-study participants constitute a distinct group, differing from the users after a full market launch. This raises important questions about the generalizability of pilot-study findings to a broader customer base and whether we can devise measures to mitigate potential pilot-study bias. To the best of our knowledge, the transportation literature has not yet addressed this research question.

2.2 Mechanisms Affecting Generalizability of Pilot-Study Findings

Various mechanisms, discussed in the research literature, can cause a pilot-study group to be non-representative. We review these mechanisms here as they help analyze our research question (see Table 1 for a summary).

A first potential mechanism is the participant's assessment of the offer, which includes general satisfaction, accuracy of fit to customer needs, and price-quality assessment. Regarding general satisfaction, the literature recognizes that satisfaction during a pilot study is inextricably linked to consumer loyalty and repurchase intention (see, e.g., Bhattacherjee, 2001; Oliver, 1999). Studies consistently show that higher customer satisfaction positively influences the intention to buy again, a finding demonstrated in contexts ranging from smartphone services to online retail (Ng et al., 2020; Bukari et al., 2025). Similarly, the accuracy of fit is crucial; when an offering aligns well with customer needs and provides superior "value-in-use", effectively "co-creating" value with customers, it enhances both satisfaction and loyalty (Bhattacherjee, 2001; Payne et al., 2008). Finally, a favorable price-quality assessment, where consumers perceive high value, also significantly increases the likelihood of repurchase (Zeithaml, 1988).

A second mechanism involves innovator traits (for a theoretical framework, see Rogers, 2003). These traits include technological readiness, trust, and media exposure. Consumers with high technological readiness are more likely to embrace innovative products early. This propensity can be measured using scales such as the updated Technology Readiness Index (TRI 2.0) (Parasuraman and Colby, 2015, see Section 5.3). In the transportation context, for example, "tech-savvy" users are more likely to be early adopters of new concepts like Mobility-as-a-Service (MaaS) (Zijlstra et al., 2020), while general ambivalence toward new technology can hinder adoption (Lv et al., 2025). Trust in the industry or provider also plays a role; individuals who trust a firm are more willing to engage in new or uncertain transactions, such as accepting a new technology (Gefen et al., 2003) or trying an untested product in a pilot study. Finally, media exposure is relevant, as early adopters typically consume more media, actively seek information about innovations and tend to be more knowledgeable about the product domain (Rogers, 2003).

A third potential mechanism influencing the decision to participate in a pilot study concerns late-adopter traits. Late adopters are often motivated by their social network rather than a desire to innovate. For instance, interpersonal influences like "contagion"—word-of-mouth or peer influence—are important for purchase decisions for new products, such as vehicles (Axsen and Kurani, 2012). Similarly, in the transportation literature, users perceive greater value in new services (like autonomous public transport) when their social network is supportive (Yuen et al.,

2023). This peer-driven adoption relates to social conformity—or, in the case of information technologies, to "social computing" (Vannoy and Palvia, 2010)—where group interaction and consensus influence individuals to adopt a product.

A fourth factor is risk aversion, defined as the preference of individuals to avoid uncertainty, i.e., choosing a product whose value is known with certainty (see again Parasuraman and Colby, 2015). Researchers study the importance of risk aversion on the adoption of new products and services in a wide array of areas, including the switching between online services (Santana and Parigi, 2015), the adoption of new transportation modes (Chorus, 2014), and the residential choices of senior citizens (Seaman et al., 2015). Typically, early adoption is negatively related to risk aversion. For instance, Yoo and Chavas (2021) demonstrate that the more risk-averse US farmers are revealed to be, the later they adopt genetically modified corn.

Fifth, price sensitivity—the degree to which demand changes in response to a price change—also plays a role. For instance, Goldsmith et al. (2005) show, in a study of Korean consumers, that innovativeness is negatively correlated with price sensitivity. Similarly, Palm (2020) find that environment-conscious customers and technophiles in Sweden adopted solar photovoltaics earlier, while later adopters were predominantly motivated by economic gains.

Sixth and finally, socio-demographic variables such as gender, age, education, and income can also help explain the diffusion of innovation (see, e.g., the meta-study by Dedehayir et al., 2017). However, the findings for which groups self-select into early product stages—and, by extension, pilot studies—are rather mixed. Income is a common exception, as it often fosters early adoption and obviously interrelates with price sensitivity.

Table 1: Potential mechanisms of a pilot-study bias

Mechanism	**Channel of mechanism (summary)**	**Literature**
Assessment of the offer of the pilot study	General satisfaction, accuracy of fit to customer needs, and a favorable price–quality assessment make customers more likely to also purchase the final product.	Zeithaml (1988); Oliver (1999); Bhattacherjee (2001); Payne et al. (2008); Ng et al. (2020); Bukari et al. (2025)
Innovator traits	Consumers with high technological readiness, trust in the industry, and greater media exposure are more likely to participate in pilot studies.	Rogers (2003); Gefen et al. (2003); Parasuraman and Colby (2015); Zijlstra et al. (2020); Lv et al. (2025)
Late-adopter traits	Individuals for whom contagion in the form of peer influence is important are less likely to participate in pilot studies.	Vannoy and Palvia (2010); Axsen and Kurani (2012); Yuen et al. (2023)
Risk aversion	Individuals who prefer to avoid uncertainty are less likely to participate in pilot studies.	Chorus (2014); Parasuraman and Colby (2015); Santana and Parigi (2015); Seaman et al. (2015); Yoo and Chavas (2021)
Price sensitivity	Pilot-study participants tend to be less price sensitive.	Goldsmith et al. (2005); Palm (2020)

3 Background

The tariff landscape within Switzerland's public-transportation network is characterized by a variety of options. On the one hand, customers who use the system infrequently can purchase single tickets, including supersaver tickets, which offer discounts of up to 70% and are mainly

available for long-distance connections.[3] Moreover, users of Switzerland's public-transportation network can purchase a Half Fare Travelcard, which allows them to pay half of the regular single-ticket price (in most cases). The Half Fare Travelcard follows a two-part tariff structure, as consumers are charged both an entry fee—CHF 190 (approximately USD 235, as of January 2026) in the first year and CHF 170 in subsequent years—and a usage fee for the single tickets (Weibel et al., 2024).

On the other hand, customers who travel frequently within Switzerland's public-transportation network can purchase unlimited-use subscriptions such as the GA Travelcard (GA). The GA can be purchased on a monthly or yearly basis and costs CHF 3,995 for one year. In addition, Switzerland has regional tariff associations that offer comparable unlimited-use subscriptions restricted to regional networks or selected tariff zones thereof.

The Half Fare Travelcard PLUS targets customers with medium-mobility demand ("below the subscription threshold"). This product, permanently introduced in December 2023, offers allowances to purchase single tickets over the course of one year. In Table 2, we present the available package options. There are three options each for adults and for youth (under 25 years). For adults, the small package costs CHF 800 and provides an allowance of CHF 1,000. The medium and large packages offer progressively higher discounts: customers pay CHF 1,500 for a CHF 2,000 allowance and CHF 2,100 for a CHF 3,000 allowance. The youth options follow a similar structure but offer even greater discounts (up to 47.5% discount on the largest package).

Notably, the product eliminates financial risk. If a customer's total spending after 12 months is less than the price they paid for the credit, they receive a refund for the difference. For instance, if a customer buys the small package (paying CHF 800) but only uses CHF 700 in tickets, they are refunded CHF 100.

[3] See Huber et al. (2022) and Thommen and Hintermann (2023) for research on the effect of supersaver tickets on customer behavior.

Table 2: Half Fare Travelcard PLUS (Final Product) and Pilot Study package options

Package option	Price (CHF)	Allowance (CHF)
Final Product		
Adults		
Small package	800	1,000
Medium package	1,500	2,000
Large package	2,100	3,000
Youth		
Small package	600	1,000
Medium package	1,125	2,000
Large package	1,575	3,000
Pilot Study		
Small package	800	1,000
Large package	2,000	3,000

Note: See https://www.sbb.ch/en/offers/half-fare-travelcard-plus (Final Product) and https://www.sbb.ch/de/kampagne/markttest-oev-guthaben.html (Pilot Study, in German), accessed on October 30, 2025.

The Half Fare Travelcard PLUS has exceeded the sales expectations of the public-transportation industry by a factor of four. By the end of 2024, after its first year on the market, 204,800 packages had been sold, indicating a slight shift from unlimited-use subscriptions to single tickets (purchased with the allowance). During 2024, GA circulation fell by 5% to 424,865. Meanwhile, the circulation of regular Half Fare Travelcards (which is a de-facto prerequisite for the new offer) rose by 6.1% to approximately 3.3 million users.[4]

Before the market launch of the Half Fare Travelcard PLUS, the Swiss public-transportation providers conducted a pilot study between December 2021 and March 2023, testing a stripped-down version. In this pilot, participants could choose between a small or a large package with allowances of CHF 1,000 and CHF 3,000, respectively, independent of the age segment. Importantly, the pilot study also differed in that the allowance could only be used for "automated ticketing", thus excluding supersaver tickets, among other tickets.

In an impact evaluation of the pilot study, Sticher and Blättler (2024) assess the treatment effect based on 911 participants, of whom 893 completed the study. Applying matching methods to compare participants with non-participants, they find that yearly expenditure of participants increased by CHF 179.7. They conclude that, accordingly, the pilot study participants exhibited a price elasticity of $|\epsilon| > 1$.

In the remainder of this article, we use specific terminology for clarity. We refer to the product launched in December 2023 as the "Final Product". In contrast, we refer to the stripped-down version tested between December 2021 and March 2023 as the "Pilot Study".

[4]See https://www.allianceswisspass.ch/de/asp/News/Newsmeldung?filterCategory=4-20&newsid=839, accessed on October 30, 2025.

4 Research Design

In this section, we first outline our overall research design in Subsection 4.1, where we describe our four-step empirical strategy. We then present the statistical frameworks used to execute this strategy. In Subsection 4.2, we describe the propensity-score matching approach, which we will use to estimate the causal effects in Steps 1 and 2. In Subsection 4.3, we then detail the models used to identify self-selection bias in Step 3 and subsequently de-bias the pilot-study results in Step 4.

4.1 Overall Research Design

To evaluate whether the behavior of the Pilot Study group differs from that of Final Product customers, we proceed in four steps. This approach, as illustrated in Figure 1, allows us to assess the generalizability of pilot-study findings.

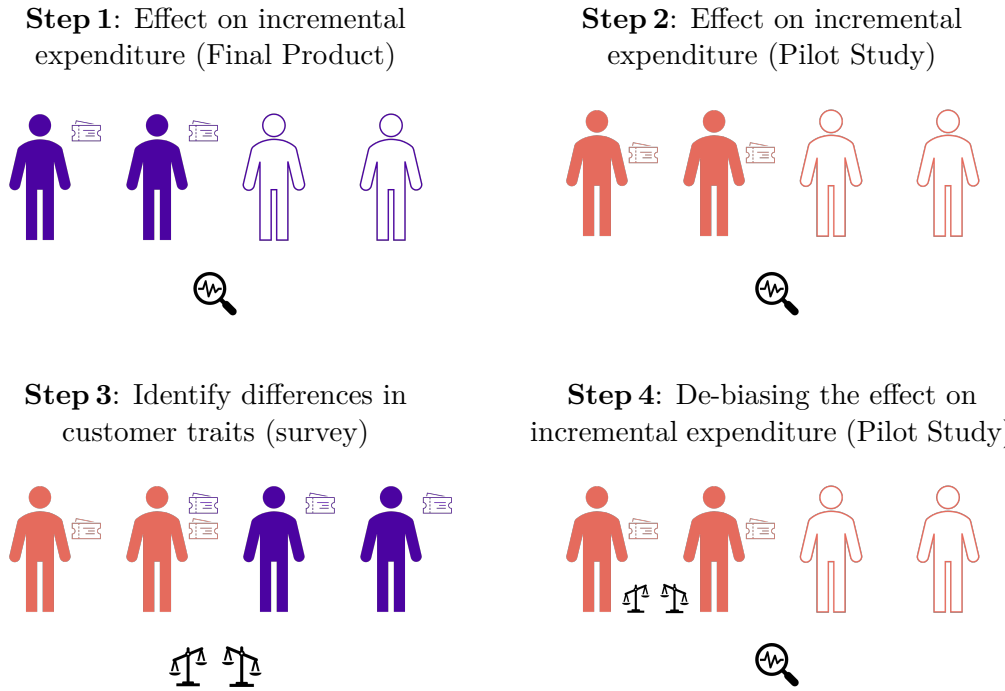


Figure 1: Overall research design

In Step 1, we assess the effect of the Final Product on expenditure. We compare randomly selected buyers with a control group of individuals who did not hold a Half Fare Travelcard PLUS and did not participate in the Pilot Study.

In Step 2, we re-assess the effect of the Pilot Study offer. We compare Pilot Study participants with a control group of individuals who were not part of the Pilot Study at that time. Note that while we closely follow the approach of Sticher and Blättler (2024), our matching method is adapted to render comparisons to the results from Step 1 feasible. Moreover, we restrict the treatment group to those individuals for whom we also have data from the following third step.

In Step 3, we survey both Pilot Study participants and Final Product customers who did not participate in the Pilot Study. Based on this survey, we compute propensity scores—the probability that an individual is a Final Product customer (represented by the scales in Figure 1). The survey is based on the mechanisms discussed in Section 2.2 that may have caused the Pilot Study group to be non-representative. In addition to calculating propensity scores, analyzing

the survey data also highlights which variables are relevant in explaining the differences between the two groups.

In Step 4, as a proof of concept, we re-estimate the incremental expenditure effect in the Pilot Study, employing odds weighting. Specifically, we weight the Pilot Study group according to the propensity scores calculated in Step 3. Comparing the results from Step 2 with those of Step 4 allows us to evaluate the success of de-biasing. Any remaining differences might be attributed to the product-design differences (e.g., package options and supersaver ticket availability) we outlined in Section 3.

4.2 Framework for Causal Effect Estimation (Steps 1 and 2)

In Steps 1 and 2, discussed in the previous subsection, we estimate the causal effect of a public-transportation policy—the introduction of the Pilot Study and Final Product offers—on customer expenditure.

Specifically, in terms of the potential outcomes framework, we are interested in the average treatment effect among the treated (ATT), i.e., the incremental effect on expenditure for the subgroup of buyers. The ATT can be formalized as

$$\text{ATT} = \mathbb{E}\left[Y(1) - Y(0) \mid D = 1\right],$$

where $Y(1)$ and $Y(0)$ denote a customer's yearly expenditure with and without treatment, respectively, and $D = 1$ when the offer is taken.

The simplest approach would be to compare mean outcomes (i.e., expenditures) between a treatment group (buyers of the Pilot Study or Final Product offers, respectively) and a control group (individuals not accepting the offer).

However, our observational data is not fully random. Participants self-select into the treatment group. Using a randomly drawn control group's average expenditure as a proxy for $\mathbb{E}[Y(0)|D = 1]$ would therefore not yield an unbiased estimate of the causal effect.

Instead, we account for systematic differences in background characteristics between the two groups. Specifically, we use these observed background characteristics $\mathbf{X}$—which influence both treatment assignment and expenditure—to build a comparable control group. Closely following Sticher and Blättler (2024), these background characteristics include: pre-treatment expenditure (total and single-ticket expenditures in the prior year); pre-treatment expenditure volatility (daily and monthly spread, daily and monthly coefficient of variation); number of trips (total and first-class); and personal characteristics (age, gender, region of residence, and customer type classification as defined by the Swiss Federal Railways).

Note that our analysis relies on the selection-on-observables assumption ($\{Y(1), Y(0)\} \perp D \mid \mathbf{X}$), which states that, conditional on $\mathbf{X}$, treatment assignment is as good as random (see, e.g., Huber, 2023).

Moreover, we rely on the common-support assumption, which requires that comparable individuals exist in both the treatment and control groups for all values of $\mathbf{X}$, which formally means that propensity scores must be strictly between zero and one ($0 < \mathbb{P}[D = 1 \mid \mathbf{X}] < 1$).

Building upon these assumptions, we employ matching, identifying pairs of individuals of the treatment and control groups that are similar in terms of background characteristics. Specifically, we perform one-to-one nearest neighbor matching (with replacement) on a Random Forest-based propensity score (based on the background characteristics described above), without a caliper.[5] That is, for each buyer we select the single best match (in terms of the propensity score) in

[5] We tuned the model's `mtry` parameter using out-of-bag (OOB) errors based on 300 trees. This yielded an optimal `mtry` of 2 for most models (3 for the adult medium package) and an OOB error rate of approximately 21.77% for the pooled analysis (on a balanced sample).

the control group, where control group individuals may be used multiple times. To account for multiple usage of control-group observations, we will use cluster-robust standard errors later on.[6]

In Table B.1 in Appendix B, we confirm the quality of the match by assessing the covariate balance before and after matching. The procedure was highly effective: the standardized mean differences for all background characteristics are negligible, and the match on the propensity score itself is very precise (the maximum distance between any pair is 0.00181). This indicates the matching successfully created a comparable control group. However, as our attention is on expenditure—and even a precise propensity-score match does not guarantee perfect balance on every single covariate—we note that small but meaningful differences in pre-treatment total expenditure remain (e.g., MU 690.48 vs. MU 678.30 when pooling packages in Step 1). We therefore additionally correct the final estimates by controlling for pre-treatment expenditure. This step effectively creates a double-robust estimator, combining matching with a "difference-in-differences" adjustment.[7]

To compute total expenditure for Final Product customers (i.e., in Step 1), we had to predict future allowance depletion as a preparatory step, since we do not observe the full behavior of all customers of the final product. We applied recursive prediction using a gradient boosting model. As predictors, we included: remaining credit, expenditure in the last 7/30 days, share of the allowance spent, month (trend), day type (workday vs. Saturday vs. Sunday), credit type (paid vs. bonus), customer age (adults vs. youth), and package size. We tuned hyperparameters using 5-fold cross-validation.[8] This approach allows us to predict the day of allowance exhaustion. (Note that for individuals in both the treatment and matched control groups, we only considered consumption within the treatment period, which we define as ending once the allowance is exhausted.)

In contrast, if we only considered "complete" observations (which we will do as a robustness check), we would encounter a two-fold bias. First, "fast-depleters" would have been overrepresented. Second, since we only observe approximately 14 months upon market launch, complete observations would have barely included any customers that start using their Half Fare Travelcard PLUS in the summer and early autumn months, where commuters (i.e., potential "down-sellers") typically consider renewing their subscriptions.

4.3 Framework for Bias Identification and Correction (Steps 3 and 4)

We perform the analysis for Steps 3 and 4 on a final analytical sample of 955 customers. We detail the derivation of this sample, including exclusion criteria and imputation procedures, in Section 5.3.

To identify the key drivers of self-selection bias (Step 3), we compare Pilot Study participants to Final Product buyers using a two-model approach. We first estimate a "flexible model" to explore all potential predictors, and from these results, we specify a final "parsimonious model" to identify the core factors.

The "flexible model" is a logistic regression where the dependent variable is an indicator for being a final-product buyer (vs. a pilot-study participant). As predictors, we include all questions on innovator traits, late-adopter traits, risk aversion, price sensitivity, as well as demographic controls. We estimate this model using Firth's penalized likelihood method (Firth,

[6] For instance, when pooling packages in Step 1, only 6,134 unique control-group individuals served as the "statistical twins" for the entire 50,000-observation treatment group, highlighting the high degree of replacement.

[7] Note that for Step 1, we also apply the difference-in-differences approach introduced by Callaway and Sant'Anna (2021) to test the sensitivity of our results. These findings are presented in Appendix A.

[8] Cross-validation yielded optimal values of `max_depth` = 5, `min_child_weight` = 5, `eta` = 0.05, and `nrounds` = 117.

1993) to account for potential (quasi-)complete separation in the data. We then assess the overall importance of each predictor using a Type II Analysis of Deviance, the logistic regression analogue of an ANOVA test.

The "parsimonious model" is a simplified logistic regression that explicitly identifies the factors differentiating the two groups, which subsequently inform the de-biasing procedure in Step 4. We specify this model by selecting only the most significant predictors from the flexible model (those statistically significant at the 95% level). When using these Likert-scale survey questions as predictors, we make the standard assumption of equal distance between scale points. To confirm the variable selection for the parsimonious model, we also employ a Random Forest classification algorithm and find that its variable importance rankings (based on Mean Decrease in Accuracy) align with our choices.[9]

Finally, to de-bias the Pilot Study results (Step 4), we use the outputs from the Step 3 models to re-weight the Pilot Study group. The goal is to adjust the sample so that it matches the factor distribution of the Final Product group. To do so, we use the odds of the propensity score (see, e.g., Westreich et al., 2017).

This method can be applied using either the flexible model or the parsimonious model. In either case, we first define the propensity score $\mathbb{P}(i = 1 \mid \mathbf{X})$ as the probability of being a Final Product buyer ($i = 1$), conditional on the set of covariates $\mathbf{X}$ used in that specific model.

We calculate this propensity for each individual using the estimated coefficients ($\hat{\boldsymbol{\beta}}$) from the chosen model in the logit formula:

$$\mathbb{P}(i = 1 \mid \mathbf{X}) = \frac{1}{1 + e^{-\hat{\boldsymbol{\beta}}' \mathbf{X}}}.$$

The odds weights applied to the Pilot Study group ($i = 0$) are then defined as:

$$w(\mathbf{X}) = \frac{\mathbb{P}(i = 1 \mid \mathbf{X})}{1 - \mathbb{P}(i = 1 \mid \mathbf{X})}.$$

Note that in the case of de-biasing based on the parsimonious model, we restrict $\mathbf{X}$ (and the corresponding estimates $\hat{\boldsymbol{\beta}}$ used) to those variables having a significant coefficient (at the 95% level), as our motivation for using a parsimonious model is, among other things, practical feasibility.

5 Data

In this section, we detail the three distinct data sources used to execute our four-step empirical strategy. We describe the administrative data for the Final Product, the data from the Pilot Study, and the supplementary survey data, respectively.

5.1 Final Product Data

We assess individual-level administrative data for buyers of the Half Fare Travelcard PLUS (the Final Product) and a control group. The Swiss Federal Railways provided data for a random sample of 50,000 Final Product customers and 50,000 non-buyers (the control group), covering the period from the product's market launch on December 13, 2023, until February 3, 2025.

For the 50,000 buyers, the dataset includes the purchase date and the size of the allowance package. In total, we observe 67,144 Half Fare Travelcard PLUS purchases within this sample, of which 60,302 were for adults and 6,842 for youth. A subset of customers purchased multiple

[9] We prioritize the Mean Decrease in Accuracy, as it is a direct measure of a variable's contribution to predictive performance, and as the Mean Decrease in Gini is known to be biased towards continuous variables such as Age, which it ranked as the most important predictor in our model.

packages during this period (most notably, 3,695 individuals purchased both the small and medium packages).

A further dataset contains detailed transaction data for both groups, covering both the post-launch period and the year prior (December 13, 2022, to February 3, 2025). This includes 14,326,242 single-ticket transactions, which we group from 458 unique ticket types into seven ticket categories.[10]

A third and final dataset includes 390,314 subscription records, including information on after-sales services (e.g., reimbursements). For the analysis, we disaggregated all ticket and subscription validity periods to the individual-day level. As we perform this disaggregation also in the pre-treatment period, and as usage can be distant in time from sale dates, the observation period for subscription is consequently more extensive (October 14, 2021 to February 2, 2025).

The resulting individual-day level data (combining single-ticket data and subscription data) has several purposes: It sums up to the total expenditure for the control group; it reveals what the allowance of the Final Product is used for; and it determines the (expected or actual) day of allowance depletion and the fraction of consumption for which a user does not pay (traveling "on bonus"). Moreover, it constitutes additional expenditure also for the treatment group (e.g., Half Fare Travelcard, regional season tickets, excluded ticket types such as campaign-related tickets or tickets for companions).

5.2 Pilot Study Data

The administrative data for the Pilot Study participants is structurally identical to that of the Final Product buyers (see Section 5.1), including detailed pre-treatment and post-treatment transaction records.

The Pilot Study, which ran from December 15, 2021, to February 23, 2023, was analyzed in Sticher and Blättler (2024). The administrative dataset for this full sample comprises 889 buyers (589 with the small package, 300 with the large package) and a control group of 15,181 individuals. This dataset includes 822,100 single-ticket transactions (December 1, 2020, to March 31, 2023) and 103,911 subscription records (October 25, 2017, to March 31, 2023).

In contrast to that study, our analyses involving the pilot focus only on the 241 Pilot Study participants from this sample who also responded to the supplementary survey (described in Section 5.3).

For the causal analysis of the Pilot Study (Step 2), we compare these 241 individuals—of whom 154 opted for the small package and 87 opted for the large package—to the original control group of 15,181 non-participants. A key methodological difference from the Final Product analysis is that we observe the complete behavior for all 241 participants, as the pilot period was concluded. This means all participants had either fully exhausted their allowances or received a final reimbursement, so no prediction of depletion was necessary.

5.3 Survey Data

In order to compare Pilot Study participants with buyers of the Final Product, we conducted a web-based survey in June and July 2025 with the support of Intervista, a market research institute. In total, 1000 individuals completed the survey. Using pseudonymized unique identifiers in the dataset, we know whether each respondent was a Pilot Study participant or a buyer of the Final Product. In Table 3, we refer to this as revealed participation. It is important to note that, for data protection reasons, we were not able to link the datasets of the Final Product to

[10] These categories are: nationwide single tickets, nationwide day passes, nationwide supersaver tickets, regional single tickets, regional day passes, regional supersaver tickets, other tickets.

those of the Pilot Study. Consequently, there is no column for "both" in the revealed participation. Furthermore, Table 3 shows that we also observe a group labeled "neither," consisting of 36 individuals for whom no consumption data are available.

Table 3: Stated vs. revealed participation

		Revealed participation		
		Pilot Study	Final Product	Neither
Stated part.	Pilot Study	58	2	1
	Final Product	60	713	34
	Both	123	8	1

* *Notes:* Number of observations: 1000.

When looking at the stated participation in the survey (Table 3), we find that of the 61 self-reported Pilot Study participants, 2 were in fact Final Product customers. Conversely, of the 807 self-reported Final Product customers, 60 were actually Pilot Study participants. This confusion is plausible given how professionally executed the Pilot Study was, as it used official distribution channels such as the regular mobile app.

Note that for our analysis, we rely on revealed participation. This principle, combined with the core goal of our weighting (representativeness), dictates how we handle the "stated both" group. Our survey data shows that customers who participated in both represent approximately two-thirds of the Pilot Study population but a negligible share of less than half a percent of the Final Product population. (This holds even under the most extreme assumption that all Pilot Study participants purchase the Final Product year-by-year; compare Section 3.)

Therefore, to create clean and representative samples for our de-biasing model, we apply the following rules: First, we retain the 123 individuals who "stated both" but were "revealed" as Pilot Study. We classify them as Pilot Study participants, in line with their revealed participation. Second, we exclude the 8 individuals who "stated both" but were "revealed" as Final Product. As we approached these individuals based on the Final Product participation—and due to the above-stated data protection measure—we cannot link their Pilot Study consumption data to their survey data.

After excluding the 36 "neither" individuals and the 8 individuals who stated "both" but were "revealed" as Final Product, we obtain an intermediate sample of 956 customers. From this sample, we conducted plausibility checks and data imputation: we removed one individual (from the Final Product group) who reported an age of 5 years (leaving 955 observations). For one individual with missing age, we imputed the median age (59 years); and for 156 individuals with missing income, we imputed values using a Random Forest classification approach.

The aim of the survey is to assess differences between pilot study participants and buyers of the final product. Therefore, the survey is based on the mechanisms summarized in Table 1 in Section 2.2. We asked respondents to rate the pilot study product, innovator traits, late-adopter traits, risk aversion, price sensitivity, and socio-demographic variables on a 6-point Likert scale (in most cases: 1 = strongly disagree, 6 = strongly agree). Table B.2 in Appendix B presents the questions asked as well as the hypotheses derived from the literature.

6 Results

In this section, we present the results of our four-step empirical strategy. We first report the causal effects of the Final Product on incremental expenditure (Step 1) and the initial Pilot Study results (Step 2). We then present the results of our selection-bias analysis (Step 3) and conclude with the de-biased Pilot Study results (Step 4).

6.1 Causal Effect of the Final Product

In Table 4, we present the estimated causal effect of the Final Product on yearly incremental expenditure. The results indicate that the effect differs widely between the heterogeneous target groups for each package. Because matching works best when studying these packages individually, we consider the weighted estimate to be the best representation of the overall average treatment effect (ATT). This effect amounts to MU 15.75 per year, approaches statistical significance at the conventional 5% level ($p = 0.06$).

Table 4: Final Product: yearly incremental expenditure

Subgroup	**Estimate**	**t-statistic**	**p-value**	**CI Lower**	**CI Upper**	**N (treated observations)**
Adults						
Small package	12.05	0.96	0.34	−12.67	36.76	38,643 (58.5%)
Medium package	33.09	2.20	0.03**	3.68	62.51	14,842 (22.5%)
Large package	22.41	0.85	0.40	−29.32	74.14	5,766 (8.7%)
Youth						
Small package	−37.85	−2.54	0.01**	−67.09	−8.61	4,748 (7.2%)
Medium package	19.34	1.48	0.14	−6.30	44.98	1,332 (2.0%)
Large package	147.29	5.69	< 0.01***	96.55	198.02	735 (1.1%)
Weighted estimate	15.75	1.85	0.06*	−0.93	32.42	66,066 (100%)
Full sample	2.90	0.23	0.82	−21.79	27.59	66,066 (100%)
Early adopters	29.01	2.89	0.04**	9.37	48.66	9,985 (15.1%)

[*] *Notes:* * p < 0.1, ** p < 0.05, *** p < 0.01. Early adopters are individuals who purchased a Half Fare Travelcard PLUS before February 3, 2024. Cluster-robust standard errors are used to account for control-group individuals being used multiple times for matching. Also note that the ATTs are weighted averages, as some individuals purchased more than one Final Product.

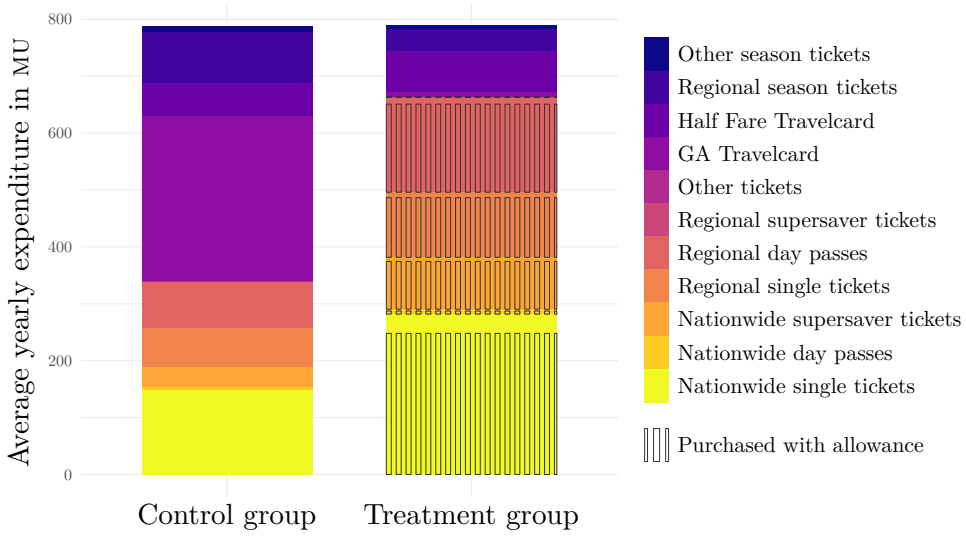
In contrast, the full-sample estimate, which pools all packages together, is smaller (MU 2.90) and not statistically significant ($p = 0.82$). While the weighted estimate is methodologically more reliable for revenue planning, this full-sample estimate is crucial for our study's main goal. To compare the Final Product with the Pilot Study in the following steps, we must rely on this "full-sample" approach for a comparable basis, because both the small sample size in the replicated Pilot Study and the overhaul in the product range prevent a more granular, package-by-package analysis.

A closer look at the package-specific ATTs in Table 4 illustrates the effect heterogeneity mentioned above. For adults, the ATTs for the small and large packages are positive—with yearly incremental expenditures of MU 12.05 and MU 22.41, respectively—but are not statistically significant. In contrast, the ATT for the medium package, amounting to MU 33.09, is statistically significant at the 5% level.

For youth, the results are even more varied. The ATT for the medium package is positive but not statistically significant. Meanwhile, the effects for the small and large packages contrast sharply. The small package corresponds to a statistically significant decrease of MU 37.85 in public-transportation expenditure per year. This negative effect suggests strong cannibalization, which may be explained by the GA Jugend (GA Youth Travelcard) offer. This nationwide subscription costs CHF 2780.– (or just CHF 710.– for families with a "baseline" GA Travelcard), making it a common alternative for this group before the Final Product's introduction. Conversely, the large package for youth corresponds to a large and highly significant increase of MU 147.29 per year.

Understanding the source of the ATT requires analyzing expenditure distribution. Figure 2 illustrates this mechanism for the Final Product by distinguishing between subscriptions and single tickets.

Figure 2: Final Product: yearly expenditure composition vs. control group



We find that the average expenditure on subscriptions (GA Travelcard, Half Fare Travelcard, regional and other season tickets) for the treatment group is about MU 321 lower than the matched control group. While expenditure on the complementary Half Fare Travelcard naturally increased, GA Travelcard cannibalization constitutes the main portion of the subscription decrease.

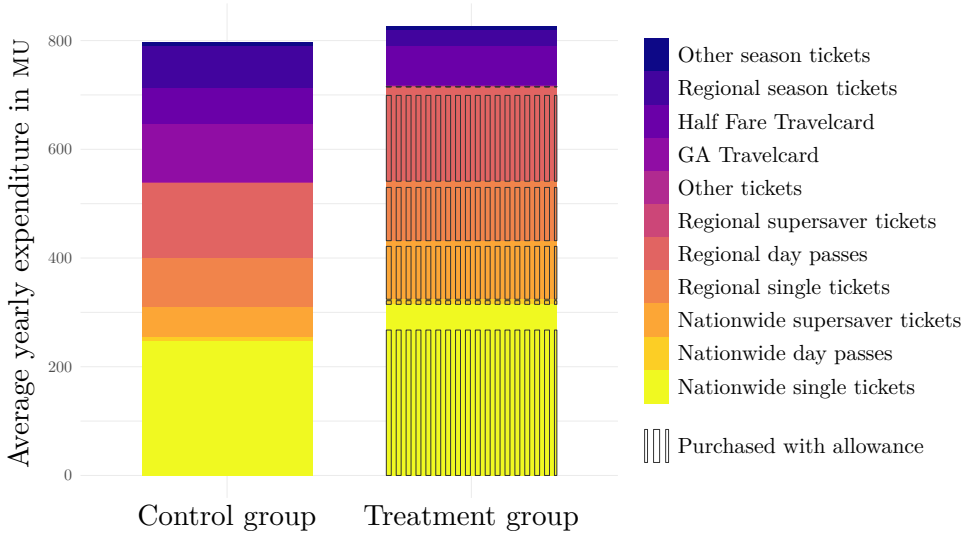
This decrease is more than offset by an approximate MU 323 increase in expenditure on single tickets. The hatched bars in Figure 2 represent expenditures made using the Final Product allowance. Not all purchases by the treatment group were made using the allowance. Besides using regional season tickets complementary to the Final Product, also single tickets (e.g., for companions) are purchased in addition to the Final Product.

Finally, we analyze the "early adopters" subgroup, which also serves as a key robustness check. As shown in Table 4, when we restrict the analysis to these customers—who purchased before February 3, 2024, and for whom no prediction of full-year behavior is necessary—the estimated effect rises to MU 29.01. The 95% confidence interval ranges from MU 9.37 to MU 48.66, indicating statistical significance.

Figure 3 reveals the mechanism for this higher effect on incremental expenditure: the cannibalization of season ticket purchases is considerably smaller for this group. This may be attributable to their different purchasing behavior, as subscription holders ("down-sellers") typ-

ically had to wait for purchasing the Final Product until their existing season tickets expired (which ofter occurs in late summer, not at the Final Product's introduction in December). "Up-sellers", i.e. existing single-ticket customers, on the other hand, have been able to jump on board immediately at market launch.

Figure 3: Final Product (early adopters): yearly expenditure composition vs. control group



This analysis of "complete" observations confirms the methodological concerns we raised in Section 4.2, where we warned that relying only on this type of observation would create a biased sample by overrepresenting "fast-depleters" and missing key commuter segments. The fact that this early-adopter ATT (MU 29.01)) is substantially different from our baseline full-sample ATT (MU 2.90) thus validates our decision to use the prediction model for the full-sample analysis. This same methodological bias is evident in the staggered difference-in-differences analysis (Appendix A), where an over-reliance on the long-term behavior of early adopters leads to a similarly inflated effect (MU 1.63 per week for the full sample).

6.2 Initial Causal Effect of the Pilot Study

Applying propensity score matching and using the sample of 241 pilot study participants who responded to the survey, we re-estimate the effect of the pilot study product on expenditure. The results indicate a mean expenditure surplus (ATT) of MU 77.47, which is statistically significant at the 5% level, as shown in Table 5. However, this initial point estimate should be interpreted with caution. The high variance of the ATT is reflected in the wide 95% confidence interval [MU 8.45, MU 112.07], suggesting that the point estimate is imprecise—likely due to the small sample size. Its consistence with the findings of Sticher and Blättler (2024) is therefore likely coincidental.

Table 5: Pilot Study: yearly incremental expenditure (initial estimate)

	Estimate	t-statistic	p-value	CI Lower	CI Upper	N (treated observations)
Full sample	77.47	2.21	0.03**	8.45	112.07	241 (100%)

* *Notes:* * p < 0.1, ** p < 0.05, *** p < 0.01. Cluster-robust standard errors are used to account for control-group individuals being used multiple times for matching. Also note that the full sample corresponds to individuals for whom survey data (Step 3) is available.

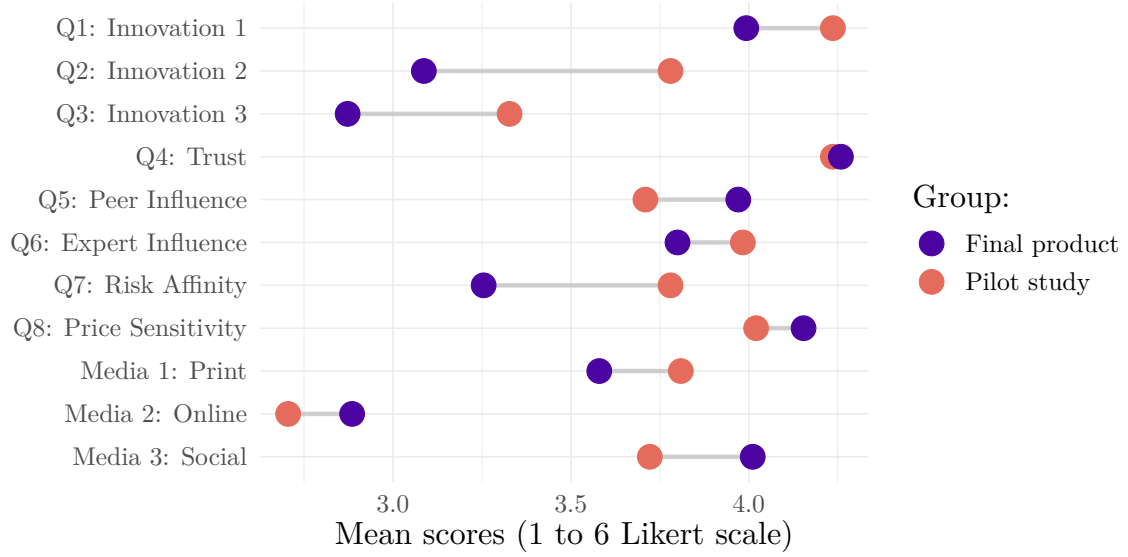
In sum, this result highlights a clear discrepancy between the Pilot Study (MU 77.47) and Final Product (MU 2.90) effect sizes, but the high imprecision of our point estimate leads us to consider the de-biasing analysis in Step 4 as a methodological proof of concept rather than a precise quantitative adjustment procedure.

6.3 Identifying Selection Bias in the Pilot Study

In this section, we compare the Pilot Study participants to individuals purchasing the Final Product (but not the Pilot Study product). This analysis proceeds in two parts: a descriptive comparison of the survey responses and a modeling procedure to identify the key drivers of selection bias, as detailed in Section 4.3.

Descriptive Analysis Figure 4 displays the average scores on the 1-to-6 Likert scale questions.[11] Regarding innovator traits, as expected, we find that consumers with higher technological readiness (according to the statements Q1 through Q3) are more likely to belong to the Pilot Study group. In contrast, trust levels vis-à-vis the public-transportation industry (Q4) are almost identical between the two groups.

Figure 4: Pilot Study vs. Final Product: survey responses



With respect to late-adopter traits, individuals for whom peer influence is important are likely to wait for the Final Product instead of participating in the Pilot Study (Q5). However, contrary

[11] Note that we were not able to ask Pilot Study assessment questions to Final Product group members. Therefore, the corresponding responses are not presented here.

to our expectations, Pilot Study participants report a higher reliance on product reviews and test reports (Q6). Accordingly, Q6 does not necessarily indicate only contagion or peer influence, but can also relate to the need for information (an innovator trait, according to Zijlstra et al., 2020).

As expected, regarding risk affinity (Q7) favors Pilot Study participation, whereas price-sensitive customers (Q8) tend to wait for the Final Product (being less experimental in their finances).

Interestingly, media exposure can indicate either Pilot Study or Final Product membership, depending on the type of the medium. While Final Product customers are more exposed to online media and social media—possibly again indicating late-adopter traits such as contagion—, Pilot Study participants more frequently consume print media, which (weakly) correlates with innovator traits (Q1 through Q3).

Finally, regarding socio-demographic characteristics, Pilot Study participants are less likely to be female (46.1% vs. 56.4% for the Final Product group), are, on average, younger (55.3 vs. 58.2 years; also see C.1 in Appendix C), more often have a bachelor's degree (55.6% vs. 48.7%), and—as expected—more often live in households with a monthly gross income above CHF 9000 (53.1% vs. 42.4%).

Flexible and Parsimonious Models Next, we present the results from our flexible and parsimonious models, described in Section 4.3. Table 6 shows the overall importance of each predictor from the Type II Analysis of Deviance for the flexible model.

We find that the self-assessment of being "often among the first to use a new technology or product" (Q2) is significant at the 1% level. In addition, on a 5% significance level, risk affinity (Q7) and consumption of print media (Media 1) also appear to be relevant factors in distinguishing between Pilot Study participants and Final Product customers.

This finding is supported by the Random Forest classification: Q2, Media 1, and Q7 are also the three most powerful predictors according to the Mean Decrease in Accuracy criterion (see Figure C.2 in Appendix C for variable importance plot).

Building upon these three selected variables (and employing the equal-distance assumption), we present the results from the parsimonious model in Table 7.

The results from this parsimonious model confirm that the self-assessment of—basically—being an innovator (Q2) is highly significant. The negative coefficient indicates that individuals with higher innovator trait values are less likely to be in the Final Product group. Similarly, risk affinity (Q7) is also still highly significant. The effect of print-media consumption, on the other hand, is not not statistically significant in this specification ($p = 0.234$). We therefore drop Media 1 from our predictor variables when de-biasing the Pilot Study incremental expenditure estimation, which we do next.

6.4 De-Biased Causal Effect of the Pilot Study

In this section, we apply the propensity-odds weighting (detailed in Section 4.3) to re-estimate the Pilot Study's causal effect. This procedure re-weights the Pilot Study sample to match the observed characteristics of the Final Product group.

Before presenting the weighted results, we assess the common support of the propensity scores. Figure 5 displays the score distributions for both the flexible and parsimonious models.

The distributions for the parsimonious model (right panel) show excellent overlap, with the Pilot Study group's scores ranging from 0.34 to 0.92 and the Final Product group's scores ranging from 0.37 to 0.96, ensuring an "apples-to-apples" comparison. In contrast, the flexible model (left panel) exhibits poorer support, with a tail of Final Product buyers having scores above 0.93 that have no counterparts in the Pilot Study group. This would lead to more volatile and less reliable

Table 6: Flexible model: analysis of deviance (type II Wald χ^2 tests)

Variable	**df**	χ^2	**p-value**
German-speaking	1	3.09	0.08*
Age	2	3.34	0.19
Female	1	2.08	0.15
Income	5	0.66	0.99
Education	6	0.82	0.99
Q1: Innovation 1	5	1.02	0.96
Q2: Innovation 2	5	15.71	0.01***
Q3: Innovation 3	5	11.05	0.05*
Q4: Trust	5	3.38	0.64
Q5: Peer Influence	5	6.16	0.29
Q6: Expert Influence	5	9.53	0.09*
Q7: Risk Affinity	5	13.26	0.02**
Q8: Price Sensitivity	5	6.86	0.23
Media 1: Print	5	14.01	0.02**
Media 2: Online	5	6.92	0.23
Media 3: Social	5	5.33	0.38

* *Notes:* Number of observations: 955. Both linear and quadratic terms were included for Age. * p < 0.1, ** p < 0.05, *** p < 0.01.

weights. Given its practical feasibility (as noted in Section 4.3) and its superior performance on the common-support assumption, we select the parsimonious model as our preferred specification for the main de-biasing analysis.

Still, we present the results from both models in Table 8 for transparency. After re-weighting, the parsimonious model yields an ATT of MU 39.59. While this point estimate is substantially lower than the initial estimate, the 95% confidence interval is wide, and the effect is not statistically significant ($p = 0.30$). The flexible model produces a similar, though slightly lower, non-significant estimate (MU 33.98).

Table 8: De-biased yearly incremental expenditure (Pilot Study group)

Model	**Estimate**	**t-statistic**	**p-value**	**CI Lower**	**CI Upper**	**N (treated observations)**
Flexible model	33.98	0.75	0.45	−54.36	122.32	241 (100%)
Parsimonious model	39.59	1.05	0.30	−34.50	113.67	241 (100%)

* *Notes:* * p < 0.1, ** p < 0.05, *** p < 0.01.

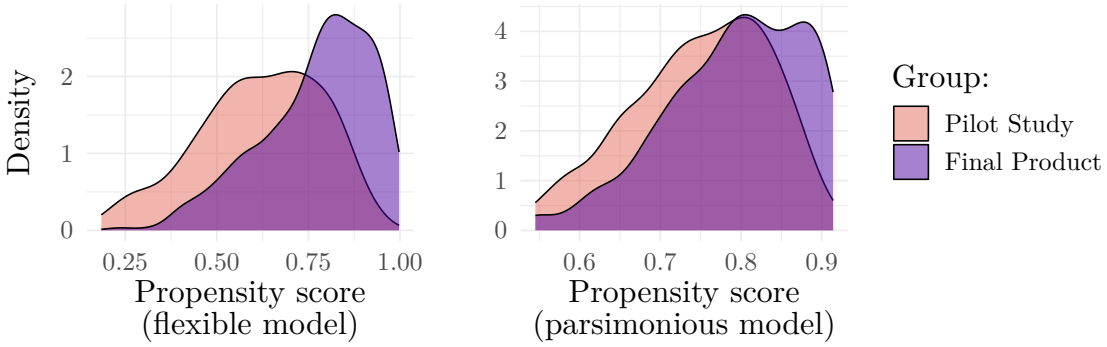
Figure 6 visualizes the impact of this de-biasing. The initial, unweighted ATT of MU 77.47 (Section 6.2) is reduced to MU 39.59. This constitutes a MU 37.89 reduction in the estimated effect, which is statistically significant (bootstrap 95% CI: [MU 3.61, MU 72.48]). This result highlights that a large portion of the initially observed "effect" was attributable to self-selection

Table 7: Parsimonious model: regression output (binary logit)

Variable	Estimate	Standard error	z-value	p-value
Q2: Innovation 2	−0.27	0.06	−4.62	< 0.001***
Q7: Risk affinity	−0.16	0.06	−2.63	0.009***
Media 1: Print	−0.05	0.05	−1.19	0.234
Constant	2.80	0.30	9.42	< 0.001***

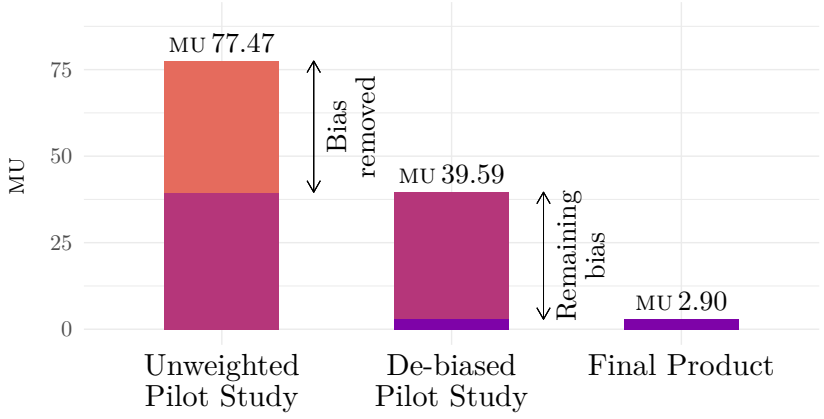
* *Notes:* Number of observations: 955. * p < 0.1, ** p < 0.05, *** p < 0.01.

Figure 5: Propensity score distributions for the pilot study and final product



bias: the Pilot Study group was, by nature, composed of individuals (e.g., innovators, risk-seekers) who were already predisposed to spend more.

Figure 6: De-biasing reduces the estimated selection bias (difference between Pilot Study and Final Product) by 51% (parsimonious model).



Nevertheless, a substantial difference remains. After correcting for self-selection, the de-biased Pilot Study effect (MU 39.59) is still much larger than the Final Product effect (MU 2.90, see Section 6.1). We interpret this remaining difference of MU 36.69 as a product effect. This

effect likely reflects the design differences between the two offers, particularly the Final Product's inclusion of more package options (e.g., the high-cannibalization youth packages) and the ability to purchase discounted supersaver tickets.

7 Discussion and Conclusion

In this study, we assessed the effect of the Half Fare Travelcard PLUS—an offer providing public-transportation credits that allow customers to purchase single tickets over the course of one year—and its corresponding Pilot Study version prior to the market launch, on customer public-transportation expenditure. The comparison between the Final Product and Pilot Study groups enables a discussion of the generalizability of pilot study findings, which is crucial for public-transportation agencies and policymakers when deciding whether to introduce new public-transportation pricing policies. Our results, based on matching methods, found the weighted average effect of the Final Product to be MU 15.75. For a direct, pooled comparison, this effect is MU 2.90, which stands in stark contrast to the MU 77.47 found in the unweighted Pilot Study group.

In addition, based on survey data, we identified key factors that explain the self-selection into the Pilot Study and Final Product groups. The relevant factors included whether a customer describes herself as "often being among the first to use a new technology or product" (Q2) or as having high risk affinity (Q7). Both of these traits made individuals significantly more likely to join the Pilot Study.

Our proof of concept then demonstrated the practical value of this finding. We showed that a simple parsimonious model, using only these two variables, was sufficient to quantify and correct for the selection bias. This correction (MU 37.89) accounted for over half of the total initial discrepancy (MU 74.58) between the two groups.

This offers a feasible method for practitioners. To improve forecasts in anticipation of market launches, public-transportation agencies could ask pilot participants these two simple questions (see Table B.2) and use the resulting scores to re-weight their sample. In our study, this is achieved by calculating the odds weights $w(\mathbf{X})$ (as defined in Section 4.3), which are derived from a propensity score $\mathbb{P}(i = 1 \mid \mathbf{X})$ based on the simple logit:

$$\hat{\boldsymbol{\beta}}'\mathbf{X} = 2.80 - 0.27 \times \text{Q2} - 0.16 \times \text{Q7}.$$

Our de-biased ATT for the Pilot Study (MU 39.59) remained significantly higher than the comparable Final Product effect (MU 2.90). We interpret this remaining MU 36.69 difference as a product effect, driven by the design changes between the two offers.

Several key differences contribute to this gap. Besides a new "medium" package, the Final Product introduced dedicated youth prices, which in the case of the small package involved cannibalization.[12]

Notably, the Final Product also included supersaver tickets, which were excluded from the Pilot Study. These tickets are a valued tool for demand management, helping to smooth peak loads and improve welfare by utilizing off-peak capacity.[13] However, they also have an estimated own-price elasticity of around -0.7 (Thommen and Hintermann, 2023). Therefore, while beneficial for shifting demand (as well as mode share and customer satisfaction), the inclusion of supersaver tickets implies a reduction in fare revenue, rendering it a prime candidate for explaining the observed product effect.

[12] Of course, instead of a short-term revenue loss, this could also be understood as a long-term investment.

[13] E.g., Huber et al. (2022) find that increasing the discount rate of supersaver tickets in Switzerland by one percentage point raises the share of rescheduled trips by 0.16 percentage points among "always buyers," i.e., customers who would have traveled even without a discount.

When it comes to limitations of our study, as repeatedly mentioned, the small number of observations that we could use for replicating and de-biasing the Pilot Study estimation (see Sections 6.2 and 6.4) renders our suggested methodology a proof of concept, which needs to be validated before relying too heavily on it. For this matter, an external validation—using another case study—would be ideal.

Another methodological limitation arises from our small sample size ($N = 241$). We had to use the same participants both to build the selection model (Step 3) and to estimate the de-biased ATT (Step 4). This procedure carries a risk of "adaptive overfitting" (see, e.g., Athey and Imbens, 2016), where the model can overfit to sample quirks, leading to extreme weights and a less stable final estimate. (While our parsimonious model's weights were moderate, with a maximum of 12.15, they still reduced the effective sample size (ESS) from 241 to 175; this was notably more stable than the flexible model's ESS of 138.) Therefore, when applying our study design in the future, we recommend splitting a (larger) sample into two parts: the first to assess differences in customer traits, and the second to estimate the effect.

Third, even after accounting for selection on early adopter traits, a violation of the selection-on-observables assumption is still conceivable. For "myRIDE," a current pilot study for yet another Swiss tariff solution, Wallimann and Sticher (2025) find that pilot-study participation is associated with key life events, such as job changes. Ignoring such events in matching creates a risk of misinterpreting "natural" expenditure fluctuations as treatment effects (both when evaluating pilot studies and final market launches). The sign of any corresponding bias is unclear, however, as these events could be associated with both "up-selling" and "down-selling" behavior.

Finally, our study design did not allow to incorporate the pilot-study participants' assessment of the offer as a mechanism for explaining the final customer structure. Other studies (see, e.g., Ng et al., 2020; Bukari et al., 2025) have shown that pilot satisfaction shapes the composition of final buyers through loyalty and repurchase intention. In our specific case study, where the 911 pilot-study participants represent a negligible fraction of the final market (over 200,000 packages sold after one year), ignoring this mechanism is likely unproblematic. In other contexts, however, we recommend that future research address this factor as well.

Despite these limitations, our findings provide a clear directive for policymakers and public-transportation agencies: Pilot-study results cannot be naively extrapolated to forecast market outcomes. We demonstrate that self-selection bias can be a substantial component of any observed discrepancy. However, we also show this bias is not an intractable problem. A parsimonious model, based on easily collected survey data on innovator traits and risk affinity, can correct for a large portion of this gap. As the public-transportation industry increasingly turns to pilot studies to test innovative pricing solutions or new mobility-as-a-service concepts, this de-biasing approach will gain in significance. Failing to account for the unique characteristics of pilot participants—the innovators and risk-seekers—will lead to biased forecasts, compromising the data-driven basis for policy and revenue planning.

References

Athey S and Imbens G (2016) Recursive partitioning for heterogeneous causal effects. *Proceedings of the National Academy of Sciences* 113(27): 7353–7360.

Axsen J and Kurani KS (2012) Interpersonal influence within car buyers' social networks: applying five perspectives to plug-in hybrid vehicle drivers. *Environment and Planning A* 44(5): 1047–1065.

Baker A, Callaway B, Cunningham S, Goodman-Bacon A and Sant'Anna PH (2025) Difference-in-differences designs: A practitioner's guide. *arXiv preprint arXiv:2503.13323* .

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

Bukari Z, Nnindini SI, Agbemabiase GC and Nyamekye MB (2025) Customer satisfaction of online purchases and repurchase intentions: the moderating role of return services. *The International Review of Retail, Distribution and Consumer Research* 35(2): 103–126.

Callaway B and Sant'Anna PH (2020) did: Difference-in-differences. URL https://bcallaway11.github.io/did/. R package version 2.1.2.

Callaway B and Sant'Anna PH (2021) Difference-in-differences with multiple time periods. *Journal of econometrics* 225(2): 200–230.

Cats O, Reimal T and Susilo Y (2014) Public transport pricing policy: Empirical evidence from a fare-free scheme in tallinn, estonia. *Transportation Research Record* 2415(1): 89–96.

Chorus CG (2014) Risk aversion, regret aversion and travel choice inertia: an experimental study. *Transportation Planning and Technology* 37(4): 321–332.

Dedehayir O, Ortt RJ, Riverola C and Miralles F (2017) Innovators and early adopters in the diffusion of innovations: A literature review. *International Journal of Innovation Management* 21(08): 1740010.

Firth D (1993) Bias reduction of maximum likelihood estimates. *Biometrika* 80(1): 27–38.

Gefen D, Karahanna E and Straub DW (2003) Trust and tam in online shopping: An integrated model. *MIS quarterly* : 51–90.

Goldsmith RE, Kim D, Flynn LR and Kim WM (2005) Price sensitivity and innovativeness for fashion among korean consumers. *The Journal of social psychology* 145(5): 501–508.

Holmgren J (2007) Meta-analysis of public transport demand. *Transportation Research Part A: Policy and Practice* 41(10): 1021–1035.

Huber M (2023) *Causal analysis: Impact evaluation and Causal Machine Learning with applications in R*. MIT Press.

Huber M, Meier J and Wallimann H (2022) Business analytics meets artificial intelligence: Assessing the demand effects of discounts on swiss train tickets. *Transportation Research Part B: Methodological* 163: 22–39.

Kębłowski W (2020) Why (not) abolish fares? exploring the global geography of fare-free public transport. *Transportation* 47(6): 2807–2835.

Kholodov Y, Jenelius E, Cats O, van Oort N, Mouter N, Cebecauer M and Vermeulen A (2021) Public transport fare elasticities from smartcard data: Evidence from a natural experiment. *Transport Policy* 105: 35–43.

Loder A, Cantner F, Adenaw L, Nachtigall N, Ziegler D, Gotzler F, Siewert MB, Wurster S, Goerg S, Lienkamp M et al. (2024) Observing germany's nationwide public transport fare policy experiment "9-euro-ticket"–empirical findings from a panel study. *Case Studies on Transport Policy* 15: 101148.

Lv X, Li J and Cai Y (2025) Unveiling consumers' ambivalence in embracing robotaxi services from a combined perspective of behavioral reasoning theory and technological readiness: a dual-stage pls-sem-ann analysis. *Current Psychology* 44(1): 371–386.

Ng H, Li T and Yeung M (2020) A pilot study of the hong kong smartphone market: How service quality affects repurchase intention via customer satisfaction and attitudinal loyalty. *International journal of business & management* 8(2): 93–109.

Oliver RL (1999) Whence consumer loyalty? *Journal of marketing* 63(4_suppl1): 33–44.

Palm A (2020) Early adopters and their motives: Differences between earlier and later adopters of residential solar photovoltaics. *Renewable and Sustainable Energy Reviews* 133: 110142.

Parasuraman A and Colby CL (2015) An updated and streamlined technology readiness index: Tri 2.0. *Journal of service research* 18(1): 59–74.

Parry IWH and Small KA (2009) Should urban transit subsidies be reduced? *American Economic Review* 99(3): 700–724.

Payne AF, Storbacka K and Frow P (2008) Managing the co-creation of value. *Journal of the academy of marketing science* 36(1): 83–96.

Redman L, Friman M, Gärling T and Hartig T (2013) Quality attributes of public transport that attract car users: A research review. *Transport policy* 25: 119–127.

Rogers E (2003) Diffusion of innovations.

Santana J and Parigi P (2015) Risk aversion and engagement in the sharing economy. *Games* 6(4): 560–573.

Schlett M and Loder A (2024) Public transport through time and space: Novel indicators for fare policy assessment. *Transportation Research Record* : 03611981251335884.

Seaman KL, Stillman CM, Howard DV and Howard Jr JH (2015) Risky decision-making is associated with residential choice in healthy older adults. *Frontiers in psychology* 6: 1192.

Sticher S and Blättler K (2024) Public-transportation credits: The potential of three-part tariffs in public transportation. *Transportation Research Part A: Policy and Practice* 182: 104022.

Thommen C and Hintermann B (2023) Price versus commitment: Managing the demand for off-peak train tickets in a field experiment. *Transportation Research Part A: Policy and Practice* 174: 103691.

Vannoy SA and Palvia P (2010) The social influence model of technology adoption. *Communications of the ACM* 53(6): 149–153.

Waldorf I, Loder A, Wurster S and Bogenberger K (2025) Low fare public transport as social policy? a study of germany's 9-euro-ticket and deutschlandticket via propensity score matching. *Transportation* : 1–41.

Wallimann H (2024) Austria's klimaticket: Assessing the short-term impact of a cheap nationwide travel pass on demand. *Transport Policy* 159: 201–214.

Wallimann H, Blättler K and von Arx W (2023) Do price reductions attract customers in urban public transport? a synthetic control approach. *Transportation Research Part A: Policy and Practice* 173: 103700.

Wallimann H and Sticher S (2025) The effect of structural changes in public-transportation tariff systems: A case study from switzerland. *Transportation Research Part A: Policy and Practice (forthcoming)* .

Weibel C, Dang L and von Arx W (2024) Choosing a flat rate over a pay-per-use tariff for public transport: Do psychological factors matter? *Transportation Research Part A: Policy and Practice* 190: 104247.

Westreich D, Edwards JK, Lesko CR, Stuart E and Cole SR (2017) Transportability of trial results using inverse odds of sampling weights. *American Journal of Epidemiology* 186(8): 1010–1014.

Yoo Di and Chavas JP (2021) An analysis of risk aversion in biotechnology adoption: the case of us genetically modified corn. *Empirical Economics* 60(5): 2613–2635.

Yuen KF, Choo LQ, Li X, Wong YD, Ma F and Wang X (2023) A theoretical investigation of user acceptance of autonomous public transport. *Transportation* 50(2): 545–569.

Zeithaml VA (1988) Consumer perceptions of price, quality, and value: a means-end model and synthesis of evidence. *Journal of marketing* 52(3): 2–22.

Zijlstra T, Durand A, Hoogendoorn-Lanser S and Harms L (2020) Early adopters of mobility-as-a-service in the netherlands. *Transport Policy* 97: 197–209.

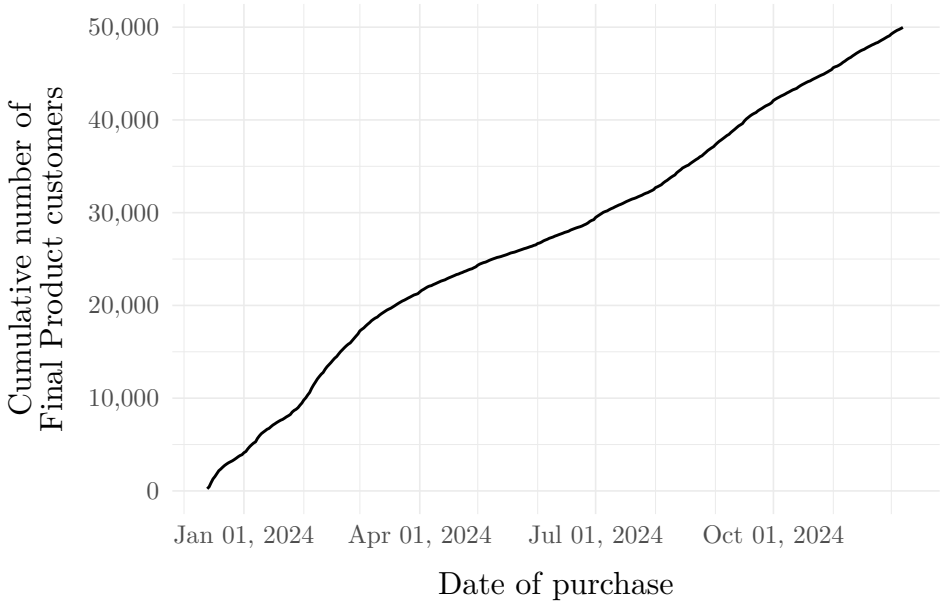
Appendices

A Causal Effect of the Final Product, Robustness Check

In Section 6.1, we present the results from applying matching to estimate the effect of the Final Product of the Half Fare Travelcard PLUS. As an alternative, we also employ a difference-in-differences (DiD) approach (see, e.g., Huber (2023); Baker et al. (2025)) to assess the effect of the Final Product on individual spending by comparing Half Fare Travelcard PLUS customers with the control group. Under the parallel-trends assumption, we estimate the average treatment effect on the treated (ATT) as the difference between the before-and-after changes among Final Product customers and the corresponding changes in the control group.

However, Figure A.1 illustrates that we face staggered treatment timing, since Final Product purchases did not all occur at the same point in time.

Figure A.1: Cumulative number of Final Product customers (sample observations) until December 2024



Therefore, in this robustness check, we follow Callaway and Sant'Anna (2021), using their `did` package in the statistical software `R` (Callaway and Sant'Anna, 2020) and construct multiple treatment subgroups of Final Product customers who purchased in the same week. We then estimate group-time treatment effects for each subgroup. Moreover, comparable to the matching approach in Section 6.1, the DiD estimator of Callaway and Sant'Anna (2021) applies propensity-score matching to calculate the individual-level group-time treatment effects, using the same covariates as in our main results. Note that, thus, our identification relies on conditional parallel trends. To obtain an overall ATT, we aggregate the subgroup-specific ATTs across different periods of exposure to the Final Product.

Again, we calculate the effect of the Final Product across all package options, i.e., including the small, medium, and large packages for both adults and youth, as well as an overall ATT obtained by merging all package options (full sample). We present the results in Table A.1.

Table A.1: Final Product: weekly incremental expenditure (robustness check)

Subgroup	**Estimate**	**t-statistic**	**p-value**	**CI Lower**	**CI Upper**	**N (treated observations)**
Adults						
Small package	0.93	2.28	0.02**	0.13	1.73	30,661 (60.7%)
Medium package	3.28	1.99	0.05**	0.05	6.51	13,295 (26.3%)
Large package	6.89	1.73	0.08*	−0.91	14.68	5,203 (10.3%)
Youth						
Small package	−0.54	−2.12	0.03**	−1.05	−0.04	4,338 (8.6%)
Medium package	−0.72	−0.83	0.40	−2.42	0.98	1,308 (2.6%)
Large package	4.99	5.90	< 0.01***	3.33	6.64	726 (1.4%)
Full sample	1.63	1.02	0.31	−1.50	4.75	50,000 (100%)

* *Notes:* * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Note that the total number of observations for individual packages exceeds the number of observations for the full sample (50,000). This is due to the fact that some customers purchased multiple (different) packages.

We see that the overall ATT amounts to an additional expenditure of MU 1.63 per week. Like in Section 6.1, however, this effect is not statistically significant at a 5% level. In contrast, the small and medium packages for adults increase weekly public-transportation expenditure by MU 0.93 and MU 3.28, respectively, both statistically significant at the 5% level (in the case of the small package, this significance is contrary to Section 6.1). Similarly, for youth, the large package shows a statistically significant additional weekly expenditure of MU 4.99. Conversely, the small youth package renders the sole statistically significant negative effect at the 5% level, amounting to MU 0.54 in weekly expenditure. Again, this result is consistent with the findings from Section 6.1.

B Additional Tables

Table B.1: Covariate balance before and after matching (Final Product, pooled across packages)

	Variable type	**Unadjusted difference**	**Adjusted difference**
Distance (propensity score)	Distance	1.700	-0.000
Customer segment 1	Binary	0.026	0.014
Customer segment 2	Binary	0.002	-0.000
Customer segment 3	Binary	0.192	-0.007
Customer segment 4	Binary	0.058	0.014
Customer segment 5	Binary	-0.073	-0.005
Customer segment 6	Binary	-0.004	-0.003
Customer segment 7	Binary	0.003	-0.000
Customer segment 8	Binary	-0.023	-0.002
Customer segment unknown	Binary	-0.180	-0.010
Age	Continuous	-0.254	-0.005
Gender	Binary	-0.007	-0.022
Region of residence	Binary	0.093	0.011
Expenditure total	Continuous	0.805	0.027
Expenditure single tickets	Continuous	0.747	0.069
Expenditure spread (daily)	Continuous	0.437	0.066
Expenditure spread (monthly)	Continuous	0.793	0.081
Expenditure CV (daily)	Continuous	0.620	0.067
Expenditure CV (monthly)	Continuous	0.010	0.026
Trips total	Continuous	0.689	0.037
Trips first class	Continuous	0.220	0.033

* *Notes:* Number of observations: 50,000; all differences are standardized using the standard deviation of the treatment group.

Table B.2: Pilot Study vs. Final Product: Mechanisms, survey statements, and hypotheses

Mechanism	**Variable name**	**Statement**	**Hypothesis (group)**
Pilot Study assessment	*Not used*	How satisfied were you overall with the public transportation credit during the market test?	Final Product (as opposed to discontinuation)
		Was the public transportation credit suitable for your personal travel behaviour?	
		How do you rate the price–performance ratio of the public transportation credit in the market test?	
Innovator traits	Q1: Innovation 1	New technologies and products contribute to a higher quality of life.	Pilot Study
	Q2: Innovation 2	I am often among the first to use a new technology or product.	Pilot Study
	Q3: Innovation 3	Friends and acquaintances come to me for advice on new technologies and products.	Pilot Study
	Q4: Trust	I trust the Swiss public transport industry to put the interests of customers first.	Pilot Study
Late-adopter traits	Q5: Peer Influence	I am guided by my social circle when trying new technologies or products.	Final Product
	Q6: Expert Influence	I rely on reviews and test reports when trying new technologies or products.	Final Product
Risk aversion	Q7: Risk Affinity	I am willing to choose riskier strategies for potentially higher rewards.	Pilot Study
Price sensitivity	Q8: Price Sensitivity	Imagine you could improve public transport. Either all connections are 10 percent faster or all prices are 10 percent lower. What do you prefer?	Final Product (for lower prices)
Media consumption	Media 1: Print	How often do you consume the following media? Printed newspapers and/or magazines	Unclear (relates to both innovator and late-adopter traits)
	Media 2: Online	How often do you consume the following media? Online media (e.g., 20 Minuten online, Watson)	
	Media 3: Social	How often do you consume the following media? Social media (e.g., Facebook, X, TikTok)	
Socio-demographics		Gender, age, highest educational attainment, income	Pilot Study (for income)

[*] *Notes:* Q1 through Q7 measured on a Likert scale from "strongly disagree" (1) to "strongly agree" (6). Q8 measured on a Likert scale from "clearly faster connections" (1) to "clearly lower prices" (6). Values for Media 1 through Media 3 are: "rarely/never" (1), "monthly" (2), "weekly" (3), "several times weekly" (4), "daily" (5), "several times daily" (6). Translations from German to English were assisted by Google Gemini.

C Additional Figures

Figure C.1: Pilot Study vs. Final Product: age distribution

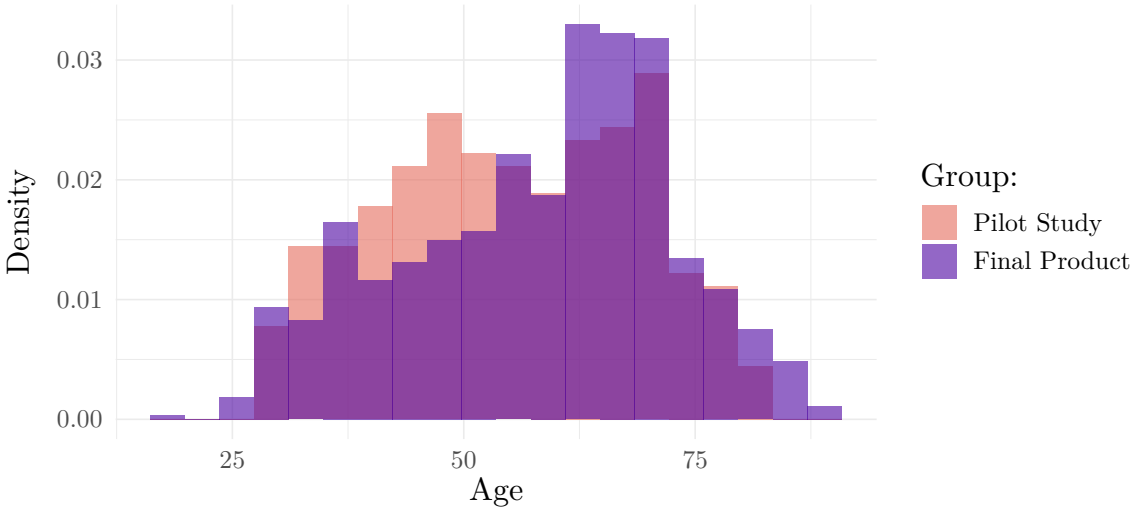
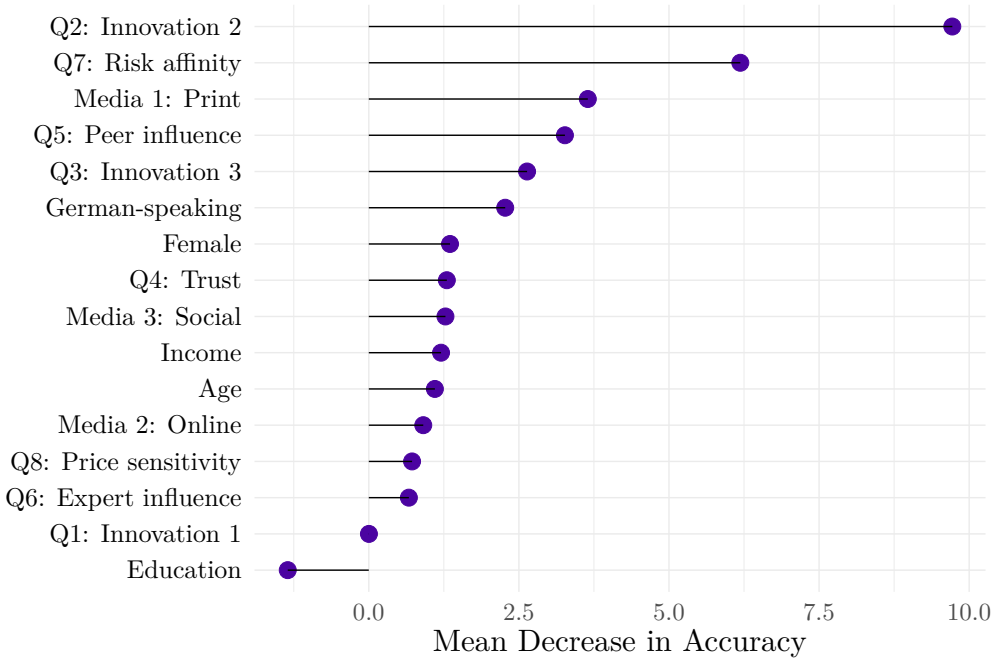



Figure C.2: Pilot Study vs. Final Product: variable importance (Random Forest classification)