The Cost of Being Late

A quantitative study examining how delivery delays affect customer purchasing behavior using transaction-level data from an online retailer

Emilio Aydin 50931

Stockholm School of Economics | BSc program in Retail Management

Bachelor Thesis NDH900 | Tutor Christoph Baldauf | May 2026

Acknowledgements

I would like to express my sincere gratitude to my supervisor, Christoph Baldauf, for valuable guidance, constructive feedback, and support throughout the thesis process. His insights have been highly appreciated and have contributed to improving the quality of this work. I would also like to thank the Center for Retailing at the Stockholm School of Economics for providing the academic setting in which this thesis was conducted. Finally, I am grateful to my family and friends for their patience and encouragement during the writing process.

Abstract

A 2022 European consumer study found that 85 percent of online shoppers say a poor delivery experience would deter them from ordering from a retailer again (Ipsos/Octopia, 2022), yet evidence on whether this stated intention translates into actual behavior remains limited to specific e-commerce contexts. Delivery delays are a recurring operational disruption in e-commerce, and prior research documents that customers who remain with a retailer after a fulfillment failure tend to spend less, wait longer between purchases, and return products more often. This study examines whether the same pattern appears in a market characterized by low switching costs and high product substitutability. Transaction-level data from a Swedish online beauty retailer covering one calendar year are used to test four behavioral hypotheses concerning customer order value, interpurchase time, product returns, and repurchase probability following a delivery delay. The results diverge from the prior literature. Customers who continued purchasing after a delivery delay did not adjust their average order value, their return rate, or their interpurchase intervals. The behavioral response detected in the data was a reduction in the probability of returning to the retailer at all. Roughly one in four delayed customers placed no further order during the observation period, and effective revenue per delayed customer fell by approximately one quarter once this silent exit was accounted for. The findings suggest that in markets with low switching costs and high product substitutability, customers do not respond to delivery failures along a continuum of reduced engagement. They respond in a binary fashion: they either tolerate the failure entirely or exit the retailer without contact. The cost of unreliable delivery in such markets is therefore concentrated in silent customer loss rather than incremental adjustment among customers who remain. This study views market structure as a prerequisite for the dynamic of gradual adjustments documented in previous research and identifies silent churn as the economically dominant consequence of delivery failures in e-commerce settings characterized by low switching costs.

Keywords: Delivery delays, Repurchase behavior, Customer retention, E-commerce, Service failure, Post-purchase disruption, Transaction data.

Table of contents

1. Introduction 3
1.1 Background 3
1.2 Purpose 6
1.3 Research Question 6
1.4 Research Contribution 6
2. Theoretical Framework 8
2.1 Expectation-disconfirmation and service failure 9
2.2 Delivery delays and customer behavior 10
2.2.1 Effects on spending 10
2.2.2 Interpurchase time 11
2.2.3 Product returns 12
2.2.4 Repurchase probability and customer churn 13
2.3 Conceptual model 15
3. Empirical Setting and Method 15
3.1 Research design 16
3.2 Empirical setting 16
3.3 Data and sample 17
3.3.1 Data cleaning 17
3.3.2 Sample overview 18
3.3.3 Before-after sample for customers experiencing a delay 21
3.4 Variable construction 22
3.4.1 Independent variable 22
3.4.2 Dependent variables 23
3.5 Empirical strategy 25
3.6 Methodological limitations 27
4. Results 27
4.1 Average order value (H1) 28
4.2 Interpurchase time (H2) 29
4.3 Product returns (H3) 30
4.4 Repurchase probability (H4) 30

4.5 Summary of results 31
4.6 Robustness checks 32
5. Discussion 36
5.1 Average order value (H1) 36
5.2 Interpurchase time (H2) 38
5.3 Product returns (H3) 39
5.4 Repurchase probability (H4) 40
5.5 Managerial implications 41
5.6 Limitations and future research 43
6. Conclusion 44
7. Reference list 45
8. Appendix 50

1. Introduction

1.1 Background

Delivery delays are a recurring operational challenge in e-commerce. A 2022 European consumer study found that 85 percent of online shoppers reported that a poor delivery experience would prevent them from ordering from that retailer again, illustrating the direct link between delivery reliability and customer retention (Ipsos/Octopia, 2022). When a promised delivery date is not met, customers experience a direct deviation from what the retailer committed to at checkout. Whether this stated propensity translates into actual behavior, and how, is the empirical question this study addresses.

In online retail, where customers cannot inspect products before purchasing and where competing retailers are easily accessible, such failures matter. Research on e-loyalty shows that customer satisfaction is a stronger driver of retention when switching costs are low (Anderson & Srinivasan, 2003). The ability to perform a promised service dependably has long been identified as the most central dimension of service quality (Parasuraman et al., 1988). A systematic review by Nguyen et al. (2018) confirms delivery timing as one of the strongest determinants of customer behavior in online retailing. Subsequent empirical work documents how fulfillment performance shapes outcomes ranging from customer satisfaction and platform ratings to repurchase and retention (Akturk et al., 2022; Harter et al., 2025). Delivery reliability is therefore not merely a logistical concern but a core component of the service a retailer offers.

These patterns are reflected in how major online retailers allocate capital. Amazon built its Prime membership around a delivery promise, expanding free next-day and same-day delivery service to over 300 million items and treating speed as a central driver of customer retention (NielsenIQ, 2024). In the Nordic market, Boozt, the region's largest online department store, has invested more than €44 million in warehouse automation at its Ängelholm fulfillment center and describes its ability to meet the delivery promise as a core competitive advantage (Scandinavian MIND, 2024). Such investments only pay off

if failing to meet a delivery commitment has meaningful behavioral consequences for customers. These investments suggest that major retailers treat delivery reliability as strategically essential. However, academic evidence on how customers react when delivery promises are not kept is still limited to specific contexts, such as quick commerce, online grocery and platform-level disruptions rather than standard e-commerce settings.

Delivery delays can arise from multiple sources within the fulfillment chain. Inventory shortages may prevent immediate dispatch, warehouse inefficiencies can slow order processing, and last-mile transportation constraints such as limited carrier capacity or routing problems may extend delivery times (Suguna et al., 2021). Although the operational causes vary, customers are usually unable to observe which part of the process failed. What they experience is simply that the promised date was not met, and it is perceived as a failure that shapes how they evaluate the retailer in the future. Research on e-commerce service failures support this view. Forbes et al. (2005) show that e-tail customers experience service failures differently than customers in traditional retail, and post-failure switching can remain high even after recovery attempts.

Prior research consistently shows that fulfillment disruptions alter subsequent customer behavior, including lower order values, reduced purchase frequency, and outright customer exit (Rao et al., 2011; Cui et al., 2019; Amorim et al., 2025). However, this evidence is drawn from contexts such as quick commerce, online grocery, and platform-level disruptions, where delivery windows and customer expectations differ from those in standard e-commerce. When customers respond to such failures by spending less, purchasing less frequently, or not returning at all, the consequences extend beyond the individual transaction and connect directly to customer retention and long-term revenue.

This study examines these dynamics using transaction-level data from Baressoshop, a family-owned Swedish online retailer operating in the beauty sector. Baressoshop sells a range of skincare, cosmetics, haircare, and fragrance products that are widely available

through direct competitors such as Lyko, Bangerhead, and Kicks. Customers face a few structural barriers to switching between retailers. The dataset comprises 40,078 orders placed by 10,938 unique customers across a single calendar year, with approximately 12.7 percent of orders arriving after the delivery date communicated to the customer at checkout. Using these records, the study examines four behavioral outcomes following a delivery delay: the value of subsequent orders (average order value), the time until the customer's next purchase (interpurchase time), the rate at which received products are returned, and the probability that the customer returns to Baressoshop at all. For a retailer competing in a market where alternatives are easily accessible, failures directly shape subsequent customer behavior. How delivery performance is managed therefore matters.

Across these four outcomes, the response to delivery delays does not follow the gradual disengagement documented in prior research. Among customers who continue purchasing, average order value and return rates show no meaningful adjustment after a delay, and interpurchase time shifts in the opposite direction to what prior work would predict. Repurchase probability, by contrast, falls by 3.67 percentage points, and 24.6 percent of delayed customers place no further order during the observation period, reducing effective revenue per delayed customer by 25.7 percent once this silent exit is accounted for. The behavioral response is binary rather than gradual: delayed customers either leave entirely or continue purchasing as before.

1.2 Purpose

The purpose of this study is to examine the behavioral consequences of delivery delays in e-commerce, specifically how delays relate to customer retention and spending patterns. By analyzing transaction-level data from Baressoshop, a Swedish online beauty retailer, the study aims to provide empirical evidence on how operational disruptions translate into measurable changes in actual customer behavior across four dimensions: order value, purchase timing, return rates and repurchase decisions.

1.3 Research Question

The study is driven by the following research question:

How do delivery delays influence customers' subsequent purchase behavior in an e-commerce context?

1.4 Research Contribution

Prior research on online order fulfillment failures consistently documents a pattern of gradual behavioral adjustment among customers who remain with a retailer after a failure. Rao et al. (2011) find that customers who experience fulfillment glitches reduce their subsequent order values; Amorim et al. (2025) report increased time to next purchase following fulfillment failures in online grocery; Harter et al. (2025) document longer interpurchase intervals after late deliveries in quick commerce. Across these settings, post-purchase disruptions are modeled as a continuum of reduced engagement, and the customer who stays after a failure is a customer who pulls back. The present study, set in a market with low switching costs and high product substitutability, finds a qualitatively different pattern and contributes to this literature in three ways.

A contradictory finding on order value. Where Rao et al. (2011) and Amorim et al. (2025) find that customers who remain after a fulfillment failure reduce their subsequent spending, the present study finds no such reduction. Among customers who continued purchasing after their first delivery delay, the change in average order value is statistically indistinguishable from zero (−24.82 SEK, $p = 0.388$). The strict-delay specification more than doubles the coefficient in magnitude (−54.73 SEK, $p = 0.136$), but the effect remains insignificant at conventional levels, and tightening the threshold further to delays exceeding five days produces a coefficient that is smaller again (−20.69 SEK, $p = 0.586$). A full-panel specification with calendar-month fixed effects yields a near-identical estimate (−23.31 SEK, $p = 0.540$). Across all three departures from the main specification, the AOV coefficient remains small in magnitude and statistically indistinguishable from zero. The absence of a spending response among retained customers is therefore not a power problem masking a small effect; it is a stable feature of

this market. This contradicts the generalized claim that customers respond to fulfillment failures by spending less and identifies market structure as a likely boundary condition.

A binary rather than a continuous behavioral response. The absence of a spending effect among retained customers, combined with a statistically significant 3.67-percentage point reduction in repurchase probability points to a behavioral pattern that prior literature has not described. Customers do not gradually disengage from the retailer after a delay; they either continue purchasing at unchanged rates or exit the retailer entirely, with no intermediate equilibrium. Of customers who experienced a delivery delay, 24.6 percent placed no further order during the observation period, and effective revenue per delayed customer falls by 25.7 percent once this exit is accounted for. Holloway and Beatty (2003) describe silent exit qualitatively as a frequent response to online service failures, and Forbes et al. (2005) document that recovery strategies often fail to prevent post-failure switching; the present study quantifies the magnitude of this exit in transaction data. It establishes that, in markets with low switching costs and high substitutability, dissatisfaction translates directly into exit rather than into observable behavioral moderation.

An interpurchase-time finding shaped by price-driven retention. Prior work (Harter et al., 2025; Amorim et al., 2025) predicts that delivery delays lengthen the time until the next purchase, as dissatisfied customers wait longer before returning. The present study finds the opposite among retained customers: interpurchase time falls by approximately four days after the first delay ($\beta = -4.17$, $p = 0.004$). This result reflects the composition of the customer base in this setting. According to information provided by the company, Baressoshop competes primarily on price and actively maintains a top position on the price-comparison platform PriceRunner, attracting a customer base whose engagement is governed by replenishment cycles and price-driven repeat buying. Customers who continue purchasing after a delivery failure are therefore drawn disproportionately from this price-sensitive segment, for whom the delivery experience is secondary to the price proposition. From a unit economics perspective, this implies that the customers retained after delivery failures may consolidate purchases less, placing orders more frequently

without a corresponding increase in order size, which raises per-shipment fulfillment cost relative to the revenue each order generates. The interpurchase-time finding, therefore, qualifies the prior literature in two ways: it shows that the predicted timing response is not universal across e-commerce contexts, and it suggests that delivery failures may shift the customer base toward less profitable segments even where per-customer change in spending is not observed.

2. Theoretical Framework

2.1 Expectation-disconfirmation and service failure

A delivery delay is a concrete deviation from what the firm committed to at checkout. From a theoretical perspective, this constitutes a service failure: performance falls short of what was expected. Oliver (1980) argues that customers evaluate outcomes relative to prior expectations, and when performance falls below those expectations, negative disconfirmation occurs, reducing satisfaction and altering subsequent behavior. A promised delivery date creates a clear reference point, and a delay shifts the outcome below that reference in a way that is directly observable to the customer. Because customers cannot see inside the fulfillment chain, a late delivery is easily attributed to the retailer rather than to external causes, amplifying the reputational consequence of the failure.

Rabinovich and Bailey (2004) confirm that delivery timeliness is a central dimension of physical distribution service quality in internet retailing, alongside availability and reliability, and that customers pay a measurable premium for superior performance across all three dimensions. Akturk et al. (2022) extend this by showing that logistics performance, including order processing time and delivery timing, is negatively associated with online customer ratings. Deshpande and Pendem (2022) document a related pattern at the customer level: even customers who were not promised fast delivery assign lower ratings once delivery times exceed two days, indicating that service quality evaluation responds to absolute delivery performance and not only to violations of an

explicit promise. Together, these findings provide a mechanism through which delivery failures can produce sustained behavioral change rather than isolated one-off reactions.

A direct empirical test of the underlying mechanism is provided by Harter et al. (2025), who use a controlled experiment in a quick commerce setting to show that customer satisfaction mediates the relationship between delivery time deviations and repurchase behavior. Their findings confirm that delivery timing affects future purchasing through its effect on how satisfied customers feel, rather than through some other channel. These perspectives establish delivery delays as a form of service failure. The mechanism runs through reduced satisfaction and damaged perceived reliability. The next section examines how this attitudinal response translates into observable changes in customer behavior.

2.2 Delivery delays and customer behavior

The studies reviewed above establish that delivery failures constitute a form of service failure that reduces satisfaction and perceived reliability. The question that follows is how this attitudinal response translates into observable behavior. Prior research documents effects across multiple distinct dimensions: how much customers spend on subsequent orders, how quickly they return, whether they return received products, and whether they come back to the retailer at all. Each outcome is reviewed below, and a hypothesis is derived from the literature.

The link between delivery performance and customer behavior in online retail has been documented across decades of research. Heim and Sinha (2001) identify delivery timeliness as one of the strongest operational drivers of customer loyalty in early electronic retailing. Subsequent empirical work has reinforced this conclusion across product categories and time periods.

2.2.1 Effects on spending

When customers perceive a retailer as unreliable, one response is to moderate their spending. A customer who experienced a failure but remains with the retailer may test it

with a smaller order before committing to larger purchases again. This cautious re-engagement logic is supported by several empirical studies. Rao et al. (2011) show that customers who experience fulfillment glitches reduce their subsequent order values, and that the magnitude of the reduction increases with the severity of the disruption. Their path model demonstrates that glitch magnitude is positively associated with customer anxiety and negatively associated with subsequent order size, suggesting that the spending reduction reflects a reallocation of value rather than a transient reaction. Amorim et al. (2025) provide further support in the context of post-purchase fulfillment failures in online grocery retail, finding that customers reduce their spending in subsequent orders following a disruption. Although their study focuses on stockouts rather than delivery delays specifically, both represent post-purchase failures that occur after the customer has already committed the transaction. Fisher et al. (2019) document the same relationship from the opposite direction, showing that rapid delivery has a positive and measurable effect on subsequent purchases in omnichannel retailing. Taken together, this evidence indicates that the spending response is bidirectional: poor fulfillment dampens subsequent order value while reliable fulfillment elevates it.

The spending response can also be understood through affective commitment. Mattila (2004) shows that service failures have a significant negative impact on customer attitudes and loyalty intentions, and that this effect is moderated by the degree of emotional bonding between customer and service provider. Customers with high affective commitment may feel betrayed by a failure, leading to a sharper attitudinal decline. Those with low commitment may respond with more muted but still negative adjustments. In an e-commerce context where affective commitment may tend to be lower than in face-to-face service settings, given the limited personal interaction between customer and retailer, this indicates that delivery failures may lead to broadly distributed, if moderate, reductions in spending across the customer base. This motivates the first hypothesis:

H1: Customers who experience a delivery delay exhibit a lower average order value on subsequent orders.

2.2.2 Interpurchase time

A second behavioral dimension on which delivery failures may register is the timing of the next purchase. If a delivery delay reduces satisfaction and perceived reliability, customers have less reason to return quickly to the retailer and more reason to explore alternatives. Harter et al. (2025) examine delivery timing in quick commerce and find that late deliveries increase the time until the next purchase. Their results also show that smaller delays have a higher marginal impact on repurchase behavior than larger ones, which indicates that frequent minor delays may be more damaging over time than occasional severe ones. Amorim et al. (2025) report a consistent pattern in online grocery retail, where post-purchase fulfillment failures increase time to the next order by approximately seven percent. Cui, Li, and Li (2019) provide causal evidence of a related dynamic: a sudden deterioration in logistics quality on a major Chinese e-commerce platform reduced platform sales by approximately fifteen percent, with the strongest effect for product categories that customers could readily source from competing platforms.

Two qualifications matter for the present study. First, these studies examine contexts, such as quick commerce, grocery retail, and platform-level disruptions, where delivery windows and customer expectations differ from those in standard e-commerce. In categories such as beauty, where customers typically expect multi-day delivery and replenishment cycles depend on product consumption rather than immediate need, the mechanism through which a delay extends interpurchase time may be weaker. Second, Rao et al. (2011) find that the reduction in purchase frequency following a fulfillment glitch is associated with elevated customer anxiety, supporting the interpretation that the timing response reflects an attitudinal rather than purely practical adjustment. On balance, prior evidence points toward a lengthening of interpurchase time following a delivery delay:

H2: Customers who experience a delivery delay exhibit a longer interpurchase time.

2.2.3 Product returns

Beyond spending and purchase timing, delivery failures may affect how customers respond to orders they do receive. A customer who has experienced a delay may approach subsequent deliveries with reduced trust in the retailer, lowering the threshold at which a product is returned. Rao et al. (2014) find that delays in the fulfillment process increase return rates, with the effect stronger for orders shipped with a promise of fast delivery than for those with standard terms. This asymmetry is consistent with the disconfirmation logic outlined in section 2.2: the higher the promised service level, the greater the disconfirmation when the promise is broken.

Beyond the immediate transaction, return behavior carries longer-term consequences for customer-retailer relationships. Petersen and Kumar (2009) show that product returns influence future purchase behavior and customer profitability, indicating that returns are not isolated events but signals that shape the broader relationship. A delivery delay that elevates returns therefore affects more than the order in question; it may alter how the customer engages with the retailer going forward.

Masuch et al. (2024) document a related but counterintuitive finding: fast deliveries, those below the average delivery time, increase return rates among new customers, which they attribute to insufficient reduction of post-purchase cognitive dissonance. Their study suggests that the relationship between delivery timing and returns is non-monotonic, and that the link between delays and returns may operate primarily through expectation-disconfirmation rather than dissonance reduction. Their findings confirm that the same mechanism connecting delivery delays to lower spending and longer interpurchase times also connects delays to elevated return behavior. A customer who received a late delivery is more likely to return the order received, consistent with a broader withdrawal of trust in the firm. At the same time, product returns in e-commerce are driven primarily by factors related to the product itself, whether the item matches the customer's expectations in terms of fit, appearance, or functional performance. Any effect of delivery timing on returns, therefore, operates on top of these product-level drivers, and its detectability may

depend on the product category. In categories where the unboxing experience and emotional context of receipt matter more, such as fashion, delivery timing has been shown to affect returns (Masuch et al., 2024). Whether the same relationship holds in categories dominated by functional product evaluation remains an open question.

H3: Customers who experience a delivery delay exhibit a higher product return rate.

2.2.4 Repurchase probability and customer churn

The behavioral adjustments discussed above, longer interpurchase times and reduced spending, represent gradual shifts in how customers engage with a retailer. However, the most consequential outcome of a delivery delay is not that customers spend less or wait longer, but that they stop purchasing altogether. When a customer does not return within a reasonable period, the retailer loses not only the next transaction, but the entire stream of future revenue that the customer relationship would have generated.

Prior research provides consistent evidence that service failures can lead to customer exit. Keaveney and Parthasarathy (2001) show that dissatisfaction is a primary predictor of switching behavior among online service customers. In their study, switchers reported significantly lower satisfaction than customers who stayed, and this relationship held even after controlling for other attitudinal and demographic factors. Forbes et al. (2005) extend this to e-commerce retail, finding that recovery strategies frequently fail to prevent switching. Their typology of e-tail failures reveals that many customers who experience service failures do not give the firm a second chance, particularly when the failure involves the core transaction, such as delivery.

Holloway and Beatty (2003) document a further pattern that is critical for interpreting retention outcomes: dissatisfied online customers often exit silently. Their qualitative study shows that many customers who experience service failures, particularly delivery-related problems, do not complain but simply stop purchasing from the retailer. Unlike a complaint, which creates an opportunity for recovery, a silent exit provides no signal until the customer is already lost. This makes churn a particularly tractable outcome in

transaction data, where exit can be inferred directly from purchasing patterns rather than from customer-initiated feedback (Tamaddoni, Stakhovych, & Ewing, 2016).

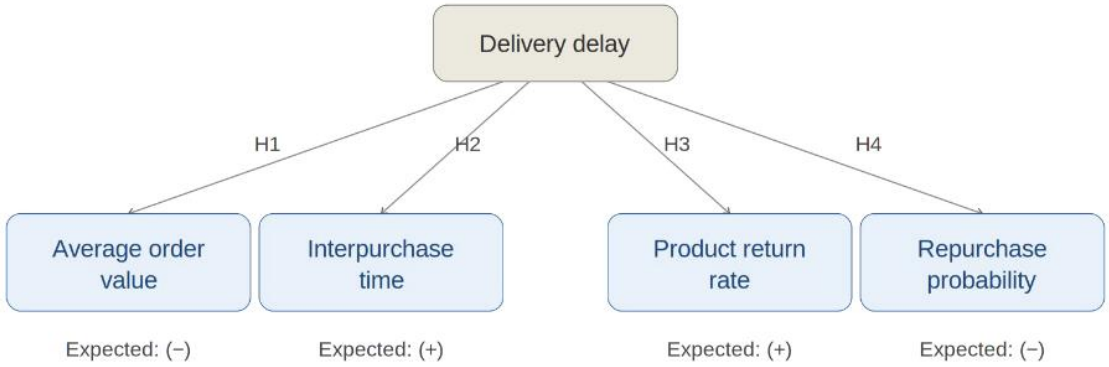
Two further factors amplify the risk of exit. Mattila (2004) shows that customers with low affective commitment to a service provider are particularly unlikely to maintain loyalty after a failure, as the emotional bond that might otherwise sustain the relationship is absent. In e-commerce, where most customer-retailer relationships are transactional, this implies that delivery failures may translate more directly into lost customers than in settings where stronger relational ties exist. Barker and Brau (2020) find that logistics disruptions reduce repurchase intentions in online retail, with stronger effects when delivery costs are high relative to order value. Amorim et al. (2025) find that fulfillment disruptions in online grocery retail not only delay the next purchase but reduce the likelihood that a subsequent order is placed at all, and that these effects occur across the customer base rather than being confined to extreme cases. The evidence points in a single direction:

H4: Customers who experience a delivery delay exhibit a lower likelihood of making a subsequent purchase.

2.3 Conceptual model

The four hypotheses capture progressively more severe consequences for the retailer: reduced revenue per transaction (H1), slower purchasing (H2), elevated returns on received products (H3), and complete loss of the customer relationship (H4). Figure 1 summarizes these reduced-form relationships.

Figure 1. Conceptual model of the relationship between delivery delays and subsequent purchasing behavior.



3. Empirical Setting and Method

3.1 Research design

This study adopts a quantitative design to examine how delivery delays relate to subsequent purchasing behavior among e-commerce customers. The analysis is grounded in archival transaction-level data capturing actual customer behavior. The unit of analysis is the individual customer, whose purchases are tracked across all transactions throughout the observation period. For customers who experienced at least one delayed delivery, the first delay serves as the focal event. Purchasing behavior prior to and including the delayed order is compared with behavior observed in the period following it, enabling within-customer comparisons that control for stable individual differences in purchase frequency and spending.

3.2 Empirical setting

The data used in this study consist of transaction-level records from Baressoshop, a Swedish online retailer operating in the beauty sector.

Customers place orders through Baressoshop's website. At the moment of purchase, all listed products appear as available. The website displays no real-time inventory levels and does not signal whether stock is constrained. The ordering process follows four steps. First, the customer selects products and proceeds to checkout. Second, at checkout, the customer is shown an expected delivery date, the delivery cost, and the return policy. These three elements together represent the retailer's commitment to the customer. Third, the customer confirms the order and receives an order confirmation by email summarizing the same information. Fourth, the order is processed in the warehouse,

packed, and handed to the carrier for shipment. According to information provided by Baressoshop, the expected delivery date is calculated from standard fulfillment and carrier lead times. The date is not revised after the order is placed, and the firm does not proactively notify customers if a delay occurs during fulfillment. Customers have access to standard order tracking and can see when the order leaves the warehouse, but they are not informed in advance if the promised date is missed. In practice, the assumption that all listed products are available is not always met. When an item turns out to be out of stock after the order is placed, Baressoshop fulfills the order by sourcing the item from an alternative supplier or by waiting for an inbound shipment, rather than cancelling the order or substituting the product. This extends the time required to ship the order and is one of the primary internal sources of delivery delays observed in the dataset. The customer does not receive notifications when this happens. The delay becomes visible only when the promised delivery date passes without the shipment arriving. The delay experience is therefore not mediated by service recovery communication from the retailer. The customer becomes aware of a delay only when the promised date is missed.

The severity of a delay is not fully captured by its absolute length. Under an expectation-disconfirmation logic, what drives dissatisfaction is the deviation relative to the reference point the retailer set at checkout. A one-day delay against a one-day delivery promise represents a 100 percent overshoot and a highly salient breach of the commitment, whereas the same one-day delay against a six-day promise is a marginal deviation that many customers may not register as a failure.

The binary delay indicator used in the main specification treats these two cases as equivalent. This is a data-driven simplification rather than a theoretical position: the dataset does not contain the promised lead time at order level in a form that allows a delay-to-promise ratio to be constructed reliably across all observations. The strict-delay robustness checks in Section 4.6 partially address severity through absolute thresholds, but a promised-lead-time-relative measure remains the theoretically preferable operationalization and is identified as a direction for future research in Section 5.6.

3.3 Data and sample

The raw dataset covers orders placed between 1 January 2025 and 31 December 2025 and includes information on order dates, delivery dates, expected delivery dates, order values, returns and unique customer identifiers.

3.3.1 Data cleaning

The dataset was pre-processed by Baressoshop before delivery to the researcher. The company removed identifying customer information and provided anonymized transaction records. On the researcher's end, cleaning was limited to the following steps. First, cancelled orders (1,280 transactions, corresponding to 3.0 percent of the raw dataset of 42,087 orders) were excluded, as a cancellation does not represent a completed purchase and would distort measures of purchase frequency and order value. The dataset does not indicate whether these cancellations were initiated by the customer or by Baressoshop. If a portion of them reflects customer-initiated cancellations in response to long expected delivery times, the behavioral consequences of poor delivery performance documented in this study may be underestimated, since such customers never enter the sample. Second, 367 records were removed due to systematic errors in the delivery date fields, such as expected delivery dates preceding the order date. These records affected 46 customer accounts. Third, an additional 362 orders lacking recorded expected delivery dates were excluded, as the delivery delay indicator could not be constructed for these observations. The resulting dataset covers orders placed between January 1 and December 31 within the observation year. The dataset does not contain orders with zero value, as these were excluded by the company before delivery.

3.3.2 Sample overview

After cleaning, the dataset has 40,078 orders placed by 10,938 unique customers. Approximately 12.7 percent of orders experienced a delivery delay. In total, 4,055 customers experienced at least one delayed delivery (37.1%), while 6,883 customers did not experience any delays during the observation period. Because each order is linked to a unique customer identifier, the data allow customer behavior to be observed across

multiple transactions over time. The distribution of delayed orders across the year is shown in Appendix 2 (Figure A2).

Customers who did not place any subsequent orders after their first delay are excluded from the average order value (H1) and interpurchase time (H2) analyses by construction but are retained in the repurchase probability analysis (H4).

Table 1. Descriptive statistics of the final dataset.

Metric	Value
Total orders	40,078
Total customers	10,938
Delay share	12.7%
Customers with delay	4,055
Customers without delay	6,883
Before-after sample (customers)	3,058

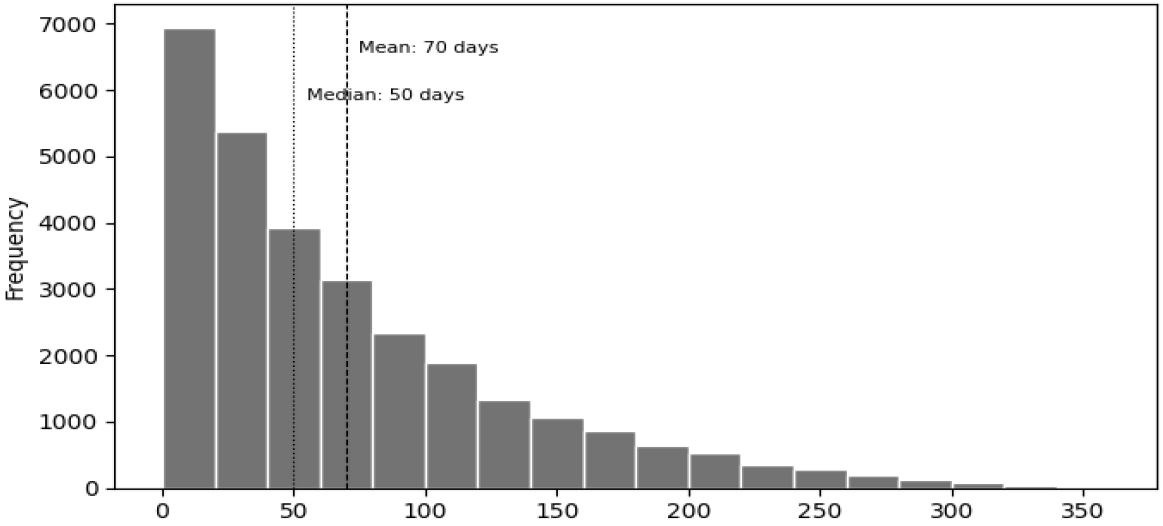
Table 2. Descriptive statistics of delivery delay.

Statistic	Value
Mean (days)	7.5
Median (days)	4.0
Standard deviation	13.1
75th percentile	9.0
90th percentile	15.0
Distribution of delay lengths	
Share of delays = 1 day	25.0%
Share of delays 2–3 days	20.4%
Share of delays 4–7 days	26.7%
Share of delays > 7 days	27.9%
Observations	**5,337**

Note. Sample restricted orders with a strictly positive delay (n = 5,337).

The 75th and 90th percentile values (9 and 15 days, respectively) are reported to characterize the upper tail of the delay distribution, where a small number of orders deviate substantially from the promised delivery date. These tail observations matter for the strict-delay robustness check reported in Section 4.6, since the behavioral response to a 15-day delay may differ qualitatively from the response to a one- or two-day delay even though both are classified as delayed under the binary indicator.

Figure 2. Distribution of interpurchase time



Distribution of interpurchase time across all customer-order pairs in the full dataset. The distribution is positively skewed, with a mean of 70 days and a median of 50 days. The H2 subsample (Section 4.2) yields different values.

Table 3. Order value statistics

Variable	All orders	Non-delayed	Delayed
Number of orders	40,078	35,007	5,071
Share of total orders	100%	87.3%	12.7%
Returned orders (N)	3,006	2,634	372
Return rate	7.5%	7.5%	7.3%

Note. Return rate is defined at the order level as the share of orders for which at least one item was returned by the customer. Return rates are similar across delayed and non-delayed orders (7.3 versus 7.5 percent); a formal test of this difference is reported in Section 4.3.

Across all 40,078 orders, the mean order value is 967.94 SEK (Swedish kronor; median 543.75 SEK; standard deviation 1,552.41 SEK), reflecting a positively skewed distribution in which a subset of high-value orders pulls the mean above the median. All monetary values reported throughout this thesis are denominated in SEK.

The magnitude of delivery delays varies considerably across affected orders. The 12.7 percent share encompasses delays of widely varying length, from one day to several weeks. Table 2 summarizes the distribution among delayed orders, computed as the difference between actual and promised delivery dates. The mean delay is 7.5 days, and the median is 4 days. A quarter of delays last a single day, while roughly 28 percent exceed one week. This dispersion supports the decision to operationalize delay as a binary indicator in the main specification, since the threshold effect of missing the promised date is uniform across delay durations. The wide variation in length, however, raises the question of whether the behavioral response varies with severity. If short delays produce minor disconfirmation that customers absorb, while longer delays cross a behavioral threshold, treating all delays as equivalent will dilute any effect concentrated among substantial failures. To examine this, a robustness specification reported in Section 4.6 restricts the delay indicator to delays exceeding three days. A meaningful change in any coefficient under this stricter definition would indicate that the behavioral response is conditional on delay severity.

3.3.3 Before-after sample for customers experiencing a delay

For the 3,058 customers who placed at least one order both before and after their first delivery delay, descriptive comparisons of order value and interpurchase time across the two periods are presented below. These customer-level descriptives motivate the regression specifications introduced in Section 3.5. The number of orders observed before and after the delay, and the share of customers with no order prior to the delay, are reported in Appendix 2 (Tables A1 and A3).

Table 4. Average order value before and after the first delivery delay

Group	Mean AOV (SEK)	N
Before delay	973.19	3,058
After delay	960.48	3,058
No-delay customers	978.94	6,883
Overall	971.21	10,984

Note. Before and after values are computed as customer-level means. N for before/after refers to the 3,058 customers in the paired sample. AOV is formally defined in Section 3.4.

The distribution of the within-customer change in average order value is reported in Appendix 2 (Figure A3).

Table 5. Interpurchase time before and after the first delivery delay

Period	Mean (days)	Median (days)	N (intervals)
Before delay	60.70	47	3,619
After delay	62.74	43	7,728

Note. Total intervals = 11,347 from 3,058 customers. Interpurchase time is formally defined in Section 3.4

Table 6. Repurchase rates by delay status

Group	Repurchase rate	N
No delay	0.888	7,215
Delay	0.852	3,351

Note. Sample restricted to customers with first focal order on or before 31 October to avoid right-censoring (n = 3,351 delayed customers in regression sample). In the full sample of 4,055 delayed customers, 24.6 percent placed no further order during the observation period.

3.4 Variable construction

3.4.1 Independent variable

The key independent variable is delivery delay, measured as a binary indicator that takes the value 1 if the actual delivery date occurs after the promised delivery date communicated to the customer at checkout, and 0 otherwise. The delay indicator was provided by Baressoshop as part of the dataset. The researcher verified that the indicator is consistent with the recorded delivery date fields, although minor discrepancies exist for a small number of orders where the actual delivery date is missing. The binary indicator flags 5,071 orders as delayed, while 5,337 orders have a strictly positive difference between actual and promised delivery dates. The 266-order discrepancy arises because the two measures are constructed from different data sources within Baressoshop's order management system. The binary indicator is set by the retailer at the point of shipment and requires both a valid promised date and a confirmed arrival date to be populated. The date difference measure, by contrast, is computed by the researcher from the raw date fields and can be calculated whenever both fields are non-missing, regardless of whether the indicator itself was logged. The 266 additional orders captured by the date-difference measure are those for which the date fields are complete, but the indicator was not populated, most often because the order was logged outside the standard fulfillment flow. The main analyses in this study use the retailer-provided binary indicator throughout, since it represents Baressoshop's own classification of which orders failed to meet the delivery promise and is the variable on which any internal performance evaluation would be based. The descriptive distribution of delay length in Table 3 is computed from the date fields directly, as delay duration cannot be derived from the indicator alone. This measure represents whether the retailer failed to meet the promised delivery time. Although the dataset contains information on delay duration, delivery delay is operationalized as a binary variable to capture the threshold effect of failing to meet the promised date. The focal delayed order in which the delivery delay occurs is included in the pre-delay period. What customers experience is whether the commitment was met, not by how many days it was missed. However, in addition to the main binary definition of delivery delay, a stricter definition (delay >3 days) is used as a robustness check.

3.4.2 Dependent variables

This study examines four measures of customers' subsequent purchasing behavior: average order value, interpurchase time, product return probability, and repurchase probability.

Average order value (AOV) is defined as the mean monetary value of orders, measured in Swedish kronor (SEK). For customers who experience a delay, AOV is calculated separately for orders placed up to and including the first delayed order and for orders placed after the delay.

Interpurchase time is defined as the number of days between two consecutive orders placed by the same customer. For customers who experience a delivery delay, interpurchase time is analyzed relative to the first delayed order, allowing comparisons before and after the delay event.

The product return rate is defined as the share of orders that are returned by the customer. It is measured at the order level as a binary indicator (1 = returned, 0 = not returned) and aggregated to represent the proportion of returned orders. For customers with a delivery delay, return behavior is analyzed relative to the first delayed order. Return rates are compared between two groups of orders for each customer. The first group consists of orders placed up to and including the first delayed order; the second consists of orders placed after this focal event. The first delayed order itself is allocated to the pre-delay group because the customer had not yet experienced the delay at the moment of placing the order; any return behavior on that order therefore reflects pre-delay trust in the retailer rather than a response to the failure. This specification permits assessment of whether experiencing a delivery delay affects the likelihood of returning items in subsequent orders. An alternative specification, which compares returns on the delayed order itself with returns on subsequent orders, would conflate the customer's pre-delay return propensity with the response to the failure and is therefore not used.

Repurchase probability is defined as the likelihood that a customer places a subsequent order. It is measured as a binary indicator equal to 1 if the customer makes another purchase, and 0 otherwise. A lack of repurchase is interpreted as customer churn. The observation period spans a single calendar year. For the repurchase analysis, orders placed after 31 October are excluded as a focal observation to reduce right-censoring of slower repurchasers. The mean interpurchase time in the dataset is approximately 60 days, but this is an average and the distribution is positively skewed, meaning that a substantial share of customers has interpurchase intervals well above 60 days. Cutting off focal observations on 31 October provides at least 60 days of look-forward time for every retained customer but does not fully eliminate censoring for customers in the upper tail of the interval distribution, who may make a subsequent purchase only after the observation window closes. Some classified non-repurchasers in this study are therefore likely to be slow returners rather than true churners. Orders placed in November and December are still observed and contribute to other parts of the analysis but are not used as focal observations in the repurchase model. The implications of the one-year observation window for the interpretation of the repurchase results are discussed further in the limitations section.

3.5 Empirical strategy

Regression models are estimated to quantify the association between delivery delays and each dependent variable while controlling for stable individual differences.

Four separate regression models are estimated, one for each dependent variable. While all models examine the relationship between delivery delays and customer behavior, the specification differs depending on the nature of the dependent variable.

H1 and H2 follow the same panel fixed effects specification:

$$Y_i t = \beta^1 \cdot AfterDelay_i t + a_i + \varepsilon_i t$$

where $Y_i t$ is the dependent variable for customer i at observation t , $AfterDelay$ a binary variable equal to 1 for observations in the post-delay period and 0 for observations in the

pre-delay period and α_i represents customer fixed effects that absorb all time-invariant differences between customers.

Order (date)	AOV	AfterDelay	Y centered	AD centered
2025-03-01	1302.29	0	−84.0	−0.5
2025-04-03 (focal)	947.60	0	−438.7	−0.5
2025-06-06	2310.69	1	+924.4	+0.5
2025-07-13	984.54	1	−401.7	+0.5

Order (date)	AOV	AfterDelay	Y centered	AD centered
2025-01-23	220.14	0	−13.1	−0.5
2025-07-24 (focal)	251.56	0	+18.3	−0.5
2025-09-15	172.13	1	−61.1	+0.5
2025-11-28	289.13	1	+55.9	+0.5

In the fixed-effects specification, α_i denotes the customer fixed effect: customer i's time-invariant level of the outcome, capturing everything stable about that customer over the observation period, such as general spending level, category preferences, price sensitivity, and baseline purchase frequency. α_i is not estimated as a reported parameter. The within (fixed-effects) estimator removes it by expressing each variable as a deviation from that customer's own mean across their orders, so the delay effect β_1 is identified exclusively from within-customer variation before and after the delay, with all stable level differences between customers absorbed.

As an illustration, customer C01031 placed two orders before the first delay (1302 and 948 SEK, the latter being the focal delayed order) and two after (2311 and 985 SEK), for a personal mean of 1386 SEK, while customer C63837 averages only 233 SEK. Expressing each order as a deviation from its own customer's mean removes α_i, so the large level gap between the two customers does not enter the estimate of β_1; only each customer's movement around their own mean is used. A single pair of customers illustrates the mechanics but does not reproduce the full-sample coefficient, which averages this within-customer movement across all 3,058 customers.

The specification includes customer fixed effects but not date fixed effects. In the current design, each customer is treated at a single point in time (their first delivery delay), which means that within a customer, the AfterDelay indicator is collinear with calendar time. Adding date fixed effects would therefore absorb much of the variation the model is designed to exploit and leave the delay effect poorly identified. A more flexible alternative would be to use the full panel and treat every delivery delay as a potential treatment event, which would allow both customer and date fixed effects to be included without absorbing the identifying variation. Such a specification would require redefining the unit of analysis, handling repeated treatment effects, and addressing the question of how customers respond to successive delays. Implementing this design is beyond the scope of the present thesis but represents the natural next step for future work.

For H3, a linear probability model is estimated at the order level, where the dependent variable indicates whether an order is returned. The model includes a delivery delay indicator and month fixed effects, with standard errors clustered at the customer level. Month fixed effects are included in the H3 specification because it is estimated across all orders rather than on before–after periods within customers, and order-level seasonality could otherwise confound the relationship between delays and returns.

For H4, a linear probability model is estimated at the customer level, examining the relationship between experiencing a delivery delay and the probability of making a subsequent purchase.

Standard errors are clustered at the customer level for the H1, H2, and H3 specifications, where multiple observations per customer create within-customer correlation in the residuals. For H4, the unit of analysis is the customer, with each customer contributing a single observation; clustering at the customer level is therefore not meaningful, and HC1-robust standard errors are reported instead to address potential heteroskedasticity in the linear probability specification.

For repurchase probability, the dependent variable is binary (0 or 1). The model is estimated using OLS, which is equivalent to a linear probability model (LPM). The LPM

is used rather than a logistic regression because the coefficient is directly interpretable as a change in the probability of repurchase, which facilitates comparison with the descriptive results.

Table 7. Model specification by hypothesis

Model	Dependent variable	Estimator	Fixed effects	Clustering	N (obs)	Customers
H1	Average order value (SEK)	Panel FE	Customer	Customer	14,405	3,058
H2	Interpurchase time (days)	Panel FE	Customer	Customer	11,347	3,058
H3	Product return (0/1)	OLS (LPM)	Month	Customer	40,078	10,938
H4	Repurchase probability (0/1)	OLS (LPM)	None	HC1 robust	10,566	10,566

3.6 Methodological limitations

The design has one limitation specific to the empirical setting that warrants flagging before the results are presented. The dataset captures transactions from a single retailer in a specific product category, which constrains the generalizability of the findings and limits what can be observed about customer behavior outside the retailer. Beauty is a category with high product substitutability and low switching costs, and the behavioral response to delivery delays observed here may differ in markets with stronger brand attachment or fewer competing retailers. The data also cannot observe what customers do outside Baressoshop, so when a customer stops purchasing after a delivery delay, the data cannot distinguish between a customer who switched to a competitor and one who left the beauty category altogether.

4. Results

Each hypothesis is tested using the specification summarized in Table 7 of Section 3.5: panel fixed-effects models for H1 and H2, and linear probability models for H3 and H4.

The four specifications are reported jointly in Table 8 below and discussed individually in Sections 4.1–4.4, following the order set out in the conceptual model.

Table 8. Regression results.

	(1) AOV	(2) IPT	(3) Returns	(4) Repurch.
Delivery delay	−24.82	−4.17***	−0.0019	−0.0367***
(SE)	*(28.78)*	*(1.46)*	*(0.0052)*	*(0.0072)*
R^2	0.0001	0.0010	0.0004	0.0027
Observations	14,405	11,347	40,078	10,566
Clusters / Customers	3,058	3,058	10,938	10,566
Customer FE	Yes	Yes	No	No
Month FE	No	No	Yes	No
Estimator	Panel FE	Panel FE	OLS (LPM)	OLS (LPM)

*Note. * p < 0.10, ** p < 0.05, *** p < 0.01. Clustered standard errors in parentheses for columns (1)–(3); HC1-robust standard errors in column (4). The dependent variable in (1) is average order value in SEK, in (2) interpurchase time in days, in (3) a binary return indicator, and in (4) a binary repurchase indicator.*

The sample size differs across specifications by design. H1 and H2 are estimated on the 3,058 customers who placed at least one order both before and after their first delay; the difference in N between the two reflects that AOV is computed at the order level while IPT is computed at the interval level. H3 is estimated at the order level across all 40,078 orders, since no before-and-after restriction is required. H4 is estimated at the customer level on the 10,566 customers whose first focal order was placed on or before 31 October, the cutoff applied to avoid right-censoring.

4.1 Average order value (H1)

H1 is tested using a panel fixed-effects specification at the customer-order level, which exploits within-customer variation in order value before and after the first delivery delay (Section 3.3.3). The descriptive comparison of customer-level means before and after the delay is reported in Section 3.3.3.

The regression yields a coefficient of −24.82 SEK on the AfterDelay indicator (SE = 28.78, p = 0.388), with a within-R^2 of effectively zero (Table 8, column 1). The point estimate is small in absolute and relative terms. Against a sample mean order value of approximately 968 SEK, a reduction of 25 SEK corresponds to about 2.6 percent of an average order. The descriptive customer-level comparison reported in Section 3.3.3 yields an even smaller decline of approximately 13 SEK, or 1.3 percent of the sample mean. Both magnitudes are statistically indistinguishable from zero. The direction is consistent with H1, but the size of the change cannot be distinguished from random variation. H1 is therefore not supported.

The coefficient of −24.82 SEK is the slope from a regression of within-customer demeaned order value on the demeaned AfterDelay indicator. Mechanically, each order's value and treatment status are expressed as deviations from that customer's own mean, which removes the customer fixed effect α_i, and β_1 is the covariance of these demeaned terms divided by the variance of the demeaned indicator. This differs from the raw before–after gap in the customer-level means reported in Table 4 (973.19 − 960.48 = −12.71 SEK) because the regression operates at the order level and weights customers by their number of orders and by the within-customer balance of pre- and post-delay observations, whereas the descriptive comparison weights every customer equally.

The null result is informative. Prior empirical work documents a spending reduction following fulfillment disruptions (Rao et al., 2011; Amorim et al., 2025), and the absence of such a reduction here is part of the pattern interpreted in Section 5.1.

4.2 Interpurchase time (H2)

H2 is tested using the same panel fixed-effects specification as H1, with interpurchase time measured in days as the dependent variable. Because customers could place multiple orders within each period, interpurchase time is computed at the interval level, with each customer serving as their own control (Section 3.5). The descriptive distribution of interpurchase intervals before and after the delay is reported in Section 3.3.3.

The regression yields a coefficient of −4.17 days on the AfterDelay indicator (SE = 1.46, p = 0.004), with a within-R^2 of 0.001 (Table 8, column 2). Against a sample mean interpurchase time of approximately 60 days, the coefficient corresponds to a shortening of about 7 percent. The coefficient is statistically significant but in the opposite direction to what H2 predicted: interpurchase time is shorter, not longer, after the first delay. The descriptive median shift from 47 to 43 days reported in Section 3.3.3 is consistent with the sign of the regression. H2 is significant in the opposite direction and therefore not supported.

The reversed direction is not a failure of detection but a substantive finding. Prior work in adjacent settings predicts that delivery failures lengthen the next-purchase interval (Harter et al., 2025; Amorim et al., 2025); the present setting yields the opposite among retained customers. The directional disagreement is interpreted in Section 5.2 and feeds into the broader contribution developed there.

4.3 Product returns (H3)

H3 is tested using a linear probability model at the order level, with a binary return indicator as the dependent variable, the delivery delay indicator as the key independent variable, and month fixed effects to absorb seasonal variation (Section 3.5). Month fixed effects are included in this specification because it is estimated across all orders rather than before/after periods for the same customer. Absorbing the month-level seasonality is therefore necessary to prevent period effects, such as holiday-season purchasing patterns, from confounding the relationship between delays and returns.

Return rates are nearly identical across the two groups: 7.52 percent for non-delayed orders and 7.34 percent for delayed orders (Table 3), with a small difference in the opposite direction to what H3 predicted.

The null result diverges from prior empirical work, which documents delivery timing effects on returns in other product categories. The substantive interpretation of this divergence is developed in Section 5.3.

4.4 Repurchase probability (H4)

H4 is tested using a linear probability model at the customer level with HC1-robust standard errors (Section 3.5). The dependent variable is a binary indicator equal to one if the customer placed any subsequent order during the observation period. Sample construction, the definition of repurchase for delayed and non-delayed customers, and the 31 October cutoff for focal observations are described in Section 3.4.2.

The regression yields a coefficient of −0.0367 on the delay indicator (SE = 0.0072, $p < 0.001$), with $R^2 = 0.003$ (Table 8, column 4). Delayed customers are 3.67 percentage points less likely to make a subsequent purchase than non-delayed customers. The non-delayed baseline repurchase rate is approximately 89 percent (Section 3.3.3). Against this baseline, the coefficient corresponds to a relative reduction of about 4 percent in the probability of returning to the retailer. H4 is supported.

Of the 4,055 customers who experienced at least one delivery delay, 997 (24.6 percent) placed no further orders during the observation period. The economic significance of this retention effect is developed in Section 5.4, where the count is interpreted as silent attrition.

4.5 Summary of results

Table 9 summarizes the results for each hypothesis. Of the four hypotheses, H4 is supported. The remaining three show responses that diverge from the predictions derived from prior literature. Delivery delays are associated with a statistically significant reduction in the probability of repurchases. The reduction in average order value (H1) is directionally consistent with expectations but not statistically significant. The change in interpurchase time (H2) is statistically significant but in the opposite direction to what was predicted. Product returns (H3) show no meaningful association with delivery delays

Table 9. Summary of hypothesis tests.

	Predicted effect	Observed effect	p-value	Result
H1	AOV decreases (−)	−24.82 SEK	0.3884	Not supported
H2	IPT increases (+)	−4.17 days	0.0042	Not supported
H3	Returns increase (+)	−0.0019	0.7143	Not supported
H4	Repurchase decreases (−)	−0.0367	0.0000	**Supported**

Note: All p-values from regression models. Observed effects report on regression coefficients.

4.6 Robustness checks

Five robustness checks are reported. The first (strict delay >3 days) tests whether the results depend on how the delay is defined. The second (60-day repurchase window) addresses the concern that the main H4 specification uses different observation windows across customers. The third (revenue incorporating churned customers) quantifies the economic magnitude of the retention effect once silent attrition is included rather than being excluded by construction. The fourth (date fixed effects) tests whether the AOV null is robust to calendar-time variation. The fifth (within-customer 60-day repurchase) tests whether the H4 retention finding is robust to a bounded observation window.

Strict delay definition (>3 days). The main specification treats any deviation from the promised delivery date as a delay, regardless of magnitude. To assess whether the results hold when attention is restricted to more substantial deviations, all four models were re-estimated using a stricter definition. A delay is coded as 1 only if the shipment arrived more than three days after the promised date. Orders with delays of one to three days are reclassified as non-delayed in this specification, which means they are pooled with on-time orders rather than excluded from the analysis. This is a deliberate choice: pooling short delays with on-time orders tests whether the behavioral response is concentrated among substantial delays, with the non-delayed group serving as a baseline against which only meaningful deviations are evaluated.

Under this stricter definition, 2,762 orders (54.5 percent of originally delayed orders) are classified as delayed, while the remaining 2,309 originally delayed orders (those with one to three days of delay) join the non-delayed group. After the reclassification, 1,814 customers have both pre-strict-delay and post-strict-delay observations and contribute to the within-customer specifications for H1 and H2. Results are reported in Table 10.

The direction of all four main coefficients is preserved under strict definition. The reduction in repurchase probability (H4) strengthens from −3.67 to −4.74 percentage points ($p < 0.001$), consistent with larger delays producing greater retention damage. The negative but imprecise AOV coefficient (H1) more than doubles in magnitude (from −24.82 to −54.73 SEK) but remains statistically insignificant at conventional levels ($p = 0.136$), indicating that a small spending effect may exist but cannot be detected with the available sample when restricted to strict delays. The interpurchase time result (H2) remains negative and significant (−5.48 days, $p = 0.003$), with the same survivorship interpretation discussed in Section 5.2 applying. The return rate coefficient (H3) shrinks further toward zero (−0.0005, $p = 0.92$), reinforcing the interpretation that delivery timing has no material effect on returns in this category.

Table 10. Main vs. strict delay specification.

Outcome	Main coef.	Main p	Strict coef.	Strict p
H1 AOV (SEK)	−24.82	0.388	−54.73	0.136
H2 IPT (days)	−4.17	0.004	−5.48	0.003
H3 Returns	−0.0019	0.714	−0.0005	0.922
H4 Repurchase	−0.0367	<0.001	−0.0474	<0.001

Note: Main H1/H2 estimated 3,058 customers with pre- and post-first-delay observations. Strict H1/H2 estimated on 1,814 customers with pre- and post-first-strict-delay observations. H3 estimated at the order level (N = 40,078). H4 estimated at the customer level with the first order/first-delay cutoff on October 31.

Stricter delay threshold (>5 days). To further test whether the absence of a spending effect reflects a power problem at moderate delay severities, the strict-delay specification was re-estimated using a more demanding threshold of more than five days. Under this definition, 2,120 orders are classified as delayed, and 1,371 customers have both pre- and post-strict-delay observations. The coefficient on AfterDelay is −20.69 SEK for AOV ($p = 0.586$) and −5.60 days for IPT ($p = 0.005$), and the repurchase probability coefficient strengthens marginally to −4.79 percentage points ($p < 0.001$). The AOV coefficient is now smaller in magnitude than under the >3-day threshold (−54.73 SEK) and remains statistically indistinguishable from zero. This pattern is informative: if a gradual spending response existed and were merely diluted by including short delays, raising the threshold should strengthen the coefficient. Instead, the coefficient becomes smaller, and the standard error rises with the smaller sample, leaving the null result intact at every severity tested.

Date fixed effects. As discussed in Section 3.5, the within-customer specification for H1 and H2 omits date fixed effects because each customer is treated at a single point in time, which makes the AfterDelay indicator collinear with calendar time. As an alternative test, the full panel is used with a time-varying treatment indicator equal to one for any order placed after the customer's first delay, allowing both customer and month fixed effects to be included. Under this specification, the AOV coefficient is −23.31 SEK ($p = 0.540$), almost identical to the main result, confirming that the null spending finding is not an artifact of unmodeled calendar-time variation. The IPT coefficient under the same specification is −33.65 days ($p < 0.001$); however, the panel-based specification identifies the effect from a different source of variation (the time-varying onset of treatment across customers rather than the within-customer before/after contrast) and is not directly comparable to the main estimate. The IPT result is reported for completeness but is not interpreted as a primary estimate.

Within-customer repurchase (60-day window). The main H4 specification compares repurchase rates between delayed and non-delayed customers at the customer level. To address the concern that the observation window differs across customers depending on

when their first delay occurred, an alternative specification estimates a within-customer panel model in which the dependent variable is a binary indicator equal to 1 if a subsequent order is placed within 60 days of the focal order. The coefficient on the AfterDelay indicator is −0.2416 (SE = 0.0083, $p < 0.001$), indicating that the probability of a repurchase within 60 days is 24.2 percentage points lower after the first delivery delay. The direction and significance are consistent with the main H4 result, although the magnitudes are not directly comparable because the two specifications measure different quantities (repurchase within the full observation window versus within a 60-day window following a given order).

Revenue incorporating churned customers. The descriptive AOV analysis in Section 4.1 compares spending only among customers who placed at least one order both before and after the first delay and therefore excludes the 997 delayed customers (24.6 percent) who never returned. To assess the economic consequence of this silent attrition, the average order value for churned customers in the post-delay period is set to zero and combined with the retained customers observed post-delay AOV. Under this specification, the effective average revenue per delayed customer falls from 974.67 SEK before the delay to 724.33 SEK after, a decline of 25.7 percent (paired $t = 10.50$, $p < 0.001$). This calculation is not a causal estimate of the delay-attributable revenue loss, as it does not adjust for the baseline churn rate observed among non-delayed customers, but it illustrates the magnitude of what the retailer effectively loses when customer exit is accounted for rather than treated as missing data.

Table 11. Summary of robustness checks.

Check	What it tests	Result	Interpretation
Strict delay (>3 days)	Whether main results hold when restricted to substantial delays	H1 (AOV): −54.73 SEK (p = 0.136); H2 (IPT): −5.48 days (p = 0.003); H3 (returns): −0.0005 (p = 0.922); H4 (repurchase): −4.74 pp (p < 0.001)	Direction preserved across all four outcomes; H4 effect strengthens vs. main spec
Stricter delay (>5 days)	Whether the AOV null reflects insufficient power at moderate delay severities	H1 (AOV): −20.69 SEK (p = 0.586); H2 (IPT): −5.60 days (p = 0.005); H3 (returns): +0.0028 (p = 0.659); H4 (repurchase): −4.79 pp (p < 0.001)	Null spending effect persists at substantial severities; H4 retention effect strengthens further
Date fixed effects (H1)	Whether the AOV null is robust to unmodelled calendar-time variation	AOV coefficient: −23.31 SEK (p = 0.540), estimated on the full panel with customer and month fixed effects	Main AOV specification confirmed; null result is not an artefact of calendar-time variation
Within-customer 60-day repurchase	Whether the H4 finding is robust to a bounded post-delay observation window	After-delay coefficient: −24.2 percentage points (p < 0.001)	Direction and significance of the main H4 result confirmed
Revenue incorporating churned customers	Economic magnitude of H4 once silent exit is included rather than excluded	Effective revenue per delayed customer: 974.67 → 724.33 SEK; −25.7% (paired t = 10.50, p < 0.001)	Silent attrition, not per-transaction adjustment, drives the economically meaningful loss

Note. Five robustness checks are reported. The first three tests whether the AOV null in the main specification is robust to alternative definitions of the treatment and to calendar-time variation. The fourth and fifth test, respectively, whether the H4 retention finding is robust to a bounded observation window and how the retention effect translates into per-customer revenue once silent attrition is included.

5. Discussion

This chapter interprets the regression results from Section 4 in relation to the prior literature reviewed in Section 2. The four behavioral outcomes are discussed in turn, beginning with an average order value. Theoretical implications are developed within each section rather than collected separately. Managerial implications follow in Section 5.5. Limitations and directions for future research are discussed in Section 5.6.

5.1 Average order value (H1)

Prior research on fulfillment disruptions documents that customers who remain after a failure reduce their subsequent spending. Rao et al. (2011) find that customers who experience fulfillment glitches place smaller orders. The magnitude of the reduction tracks the severity of the disruption. Amorim et al. (2025) report a consistent pattern in

online grocery, where post-purchase fulfillment failures are followed by lower spending on the next order. Both studies position the spending response as cautious re-engagement: the customer remains with the retailer but tests it with a smaller order before recommitting.

The present study does not find this pattern. The within-customer panel fixed-effects of regression yield a coefficient of −24.82 SEK on the AfterDelay indicator. The strict-delay specification more than doubles the magnitude to −54.73 SEK, but the result remains insignificant. Customers who continue purchasing after a delivery delay do not adjust their order value in a statistically detectable way. The robustness of this null is itself a part of the contribution. Tightening the delay threshold from any delay to more than three days, and then to more than five days, produces no detectable spending response. Replacing the within-customer design with a full-panel specification with calendar-month fixed effects yields the same null. If a small gradual spending response were operating beneath the aggregate, at least one of these specifications would be expected to reveal it. None do. The cautious re-engagement dynamic does not arise in this market not because the empirical design lacks power, but because the underlying behavior is absent.

The divergence from prior work is the contribution. Rao et al. and Amorim et al. observe their effects in settings where customers have structural reasons to remain partially engaged. In Rao et al., this engagement comes from prior purchase history and product loyalty. In Amorim et al., it comes from stockpiling, store loyalty programs, and routine grocery shopping. Baressoshop operates in a different setting. It competes against multiple substitutable alternatives on a price-comparison platform, and these structural reasons to stay are absent. The customer who continues purchasing after a delay does so because the price and assortment continue to serve them. There is no motivation to test the retailer with a smaller order. The customer has already decided that the retailer remains acceptable on the dimensions that matter most. The cautious re-engagement dynamic identified in prior work is therefore not absent because the data cannot detect it. It does not arise because the market structure does not require it.

5.2 Interpurchase time (H2)

Harter et al. (2025) find that delivery delays in quick commerce lengthen the time until the next purchase. Smaller delays produce a higher marginal impact on repurchase timing than larger ones. Amorim et al. (2025) report a related pattern in online grocery, where post-purchase failures extend the next-purchase interval by approximately seven percent.

The underlying logic in both studies is that dissatisfaction makes customers slower to return to the retailer. The present study finds the opposite among customers who return after a delay. The regression coefficient on AfterDelay is −4.17 days. Interpurchase time falls rather than rises. The result is statistically significant but in the opposite direction to what H2 predicted.

The reversal reflects a two-step selection dynamic. First, the H2 sample is restricted to the 3,058 customers who placed at least one order both before and after their first delay; the 997 delayed customers who never returned are excluded by construction. If the customers most negatively affected by the delay are the customers who exit (as the H4 result suggests), the retained sample is biased toward customers who tolerated the failure. Second, this retained sample is not a random subset of delayed customers. According to information provided by Baressoshop, the company competes primarily on price and actively maintains a top position on the price-comparison platform PriceRunner. The customers who continue purchasing after a delivery failure are therefore drawn disproportionately from a price-sensitive segment whose purchase behavior is governed by replenishment cycles and price search rather than by service quality evaluation. For such customers, the trigger for the next purchase is consumption rather than satisfaction with the previous order. A delivery delay does not change when a bottle of moisturizer is finished. The two mechanisms compound: survivorship determines who remains in the analytical sample, and the price-driven positioning of the retailer determines what that sample's behavior looks like. The shorter interpurchase time observed after a delay is the joint product of both.

This has implications that are not visible in the headline coefficient. Smaller and more frequent orders raise the per-shipment fulfillment cost without offsetting revenue. If delivery failures shift the retained customer base toward the price-sensitive segment, the operational consequences may be larger than the absence of a negative timing effect suggests. The predicted timing response documented in Harter et al. and Amorim et al. is therefore not universal across e-commerce contexts. A result that looks favorable at the level of the regression coefficient can mask compositional shifts that degrade unit economics.

5.3 Product returns (H3)

Rao et al. (2014) find that fulfillment delays increase return rates. The effect is stronger for orders shipped with a promise of fast delivery, rather than for orders with standard terms. Masuch et al. (2024) document a related finding in online fashion. Fast deliveries below the average delivery time also increase return rates among new customers. The relationship between delivery timing and returns is therefore non-monotonic and operates through expectation-disconfirmation.

The present study finds no association between delivery delays and product returns. The order-level linear probability model yields a coefficient of −0.0019. Return rates are 7.3 percent among delayed orders and 7.5 percent among non-delayed orders. The two groups are statistically indistinguishable. The strict-delay specification reinforces the null, with a coefficient of −0.0005.

The divergence reflects the category. Returns in beauty are driven by product-level factors: whether a skincare product causes an adverse reaction, whether a shade matches the customer's expectations on application, or whether a fragrance appeals on first use. These evaluations happen after the customer has opened and tried the product. They are largely independent of how the order was delivered. Rao et al. and Masuch et al. both work in product categories where the unboxing context and the emotional experience of receipt matter more. Fashion is the clearest example. Fit uncertainty and visual

expectation interact with the experience of opening the package. The contribution here is to identify a category boundary on the delivery-return relationship. Where product evaluation is primarily functional, returns operate independently of delivery timing. Retailers in such categories should not expect investments in delivery reliability to translate into lower return rates. Return data cannot be used as a leading indicator of delivery problems.

5.4 Repurchase probability (H4)

Holloway and Beatty (2003) describe silent exit as a frequent response to online service failures. In their qualitative study, customers who experienced delivery problems often did not complain. They simply stopped purchasing. Forbes et al. (2005) extend this in their typology of e-commerce service failures. Recovery strategies frequently fail to prevent switching when the failure involves the core transaction, such as delivery. Both studies establish the phenomenon. Neither quantifies its magnitude in transaction data.

The present study finds that delayed customers are 3.67 percentage points less likely to repurchase than non-delayed customers. Of the 4,055 customers who experienced at least one delivery delay, 997 (24.6 percent) placed no further orders during the observation period. The strict-delay specification strengthens the effect to 4.74 percentage points. Longer delays therefore produce larger retention consequences. To translate the per-customer effect into an order of magnitude, the post-delay decline of roughly 250 SEK in effective revenue per delayed customer, applied across the 4,055 customers who experienced at least one delay, implies a revenue gap on the order of one million SEK over the observation window. This figure is more illustrative than causal, it does not net out the baseline churn observed among non-delayed customers, but it makes the central point concrete. This aggregate loss is concentrated almost entirely in the silent-exit segment rather than distributed across the retained customer base. The 997 customers who never returned contribute zero post-delay revenue by construction, while the 3,058 customers who continued purchasing show an AOV change that is statistically indistinguishable from zero (Section 4.1). The arithmetic of the revenue decline therefore traces almost wholly to exit rather than to marginal adjustment among retained customers, which is the empirical signature of the binary response argued throughout this discussion.

Set against the modest per-shipment cost of a single late delivery (carrier penalties, service contact, compensation), the dominant cost of unreliable delivery in this market is the foregone lifetime value of silently exiting customers, which never appears in shipment-level accounting.

The contribution is the quantification of silent attrition in transaction data. The qualitative pattern documented by Holloway and Beatty and by Forbes et al. now has a magnitude attached to it. In markets like the one studied, roughly one in four customers who experience a delivery delay disappear from the retailer's records. The behavioral interpretation is consistent with the prior literature, but the economic visibility of the effect is new. The exit happens silently, with no complaint or feedback signal. The retailer has no opportunity for service recovery. The cost is invisible at the level of individual transactions but substantial when aggregated.

This result anchors the interpretation of the previous three sections. The absence of a spending effect (5.1), the shorter rather than longer interpurchase time (5.2), and the unchanged return rate (5.3) are not evidence that delivery delays carry no behavioral consequence. They are evidence that the consequence is concentrated in the exit. The customers who would otherwise have produced detectable downward shifts in spending or upward shifts in returns are the customers who have already left the sample. What the H1, H2 and H3 regressions capture is the behavior of customers for whom the delay did not exceed their tolerance threshold.

Taken together, the four results combine into a single theoretical claim foreshadowed in Section 1.4: in a setting characterized by low switching costs and high product substitutability, customers do not respond to delivery failures along a continuum of reduced engagement. They respond in binary fashion. Those for whom the failure exceeds a tolerance threshold leave entirely. Those who continue to purchase do so at that rates remain unchanged compared to before the delay. The intermediate equilibrium documented by Rao et al. (2011), Amorim et al. (2025), and Harter et al. (2025), in which

the customer stays but pulls back, does not appear in this dataset. As discussed in Section 1.1, Anderson and Srinivasan (2003) argue that satisfaction operates as a stronger driver of loyalty when switching costs are low. The binary response observed here is consistent with that logic taken in its most extreme form. When switching is costless and substitutes are direct, satisfaction operates as a discrete rather than continuous determinant of retention.

5.5 Managerial implications

Three implications follow for retailers operating in competitive, low-switching-cost environments. Each is grounded in a specific empirical result and addresses a decision that retailers in such markets actually face.

The first concerns how delivery reliability is evaluated financially. Per-shipment cost analysis treats delayed delivery as a contained operational failure, with a known cost associated with carrier penalties, customer service contact, or compensation. The 24.6 percent silent attrition rate documented in Section 5.4 implies a different unit of analysis. The relevant cost is not the shipment, but the customer's lifetime value is lost when a delay triggers exit. The exit response is concentrated in customers who never appear in complaints or feedback data. The retailer cannot recover the lost value through service recovery. The investment case for delivery reliability is therefore stronger than per-shipment economics imply. The break-even point on investments in carrier upgrades, warehouse automation, or capacity expansion is reached sooner than a per-shipment analysis would predict.

The second concerns recovery system design. Reactive recovery responds to customer-initiated contact. It cannot reach the customers at greatest risk because those customers leave silently. The 24.6 percent figure is itself the population that a complaint-driven recovery system never sees. Proactive recovery is different. Automated notifications, apology messages, or compensation are triggered by the retailer's own detection of a delay rather than by the customer's complaint. This is the mechanism that can intervene

before exit. The cost of operating proactive recovery must be weighed against the cost of the exits it prevents. The present results help quantify that trade-off by identifying that exit, rather than gradual disengagement, is the relevant outcome to prevent.

The third concerns the composition of the retained customer base over time. If repeated delivery failures gradually shift the customer base toward price-sensitive segments, as it is consistent with the results in Sections 5.1 and 5.2. The customer mix itself becomes less profitable. Retailers may continue to observe stable order values and return rates while the customers placing those orders consolidate purchases less and respond more sharply to small price differences from competitors. Delivery reliability functions both as a customer retention lever and as a means of preserving the profitability mix of the customer base. Retailers should track not only how many customers stay after delivery failures but also how their composition changes. Headline retention numbers can obscure unfavorable shifts in segment mix.

5.6 Limitations and future research

The findings should be interpreted with four caveats. Each suggests a direction for future research.

First, the observation window spans a single calendar year. This produces two related constraints on the interpretation. The first concerns the persistence of the exit effect: customers classified as silent exits in this study may eventually return, and customers classified as repurchasers may exit later. The 24.6 percent silent attrition figure therefore captures medium-term exit rather than permanent churn. The second concerns the identification of the focal delay. The within-customer before-and-after design treats the first delay observed in the dataset as the customer's first delivery delay, but customers who joined Baressoshop before the observation window may have experienced earlier delays that are not visible in the data. For these customers, the pre-delay period analyzed here may already contain post-delay behavior from prior unrecorded failures, which would attenuate any before-after comparison toward zero. A panel of two to three years would address both constraints. It would allow the persistence of the exit effect to be estimated, and it would push the unobserved-history problem further back in time so that

more of the dataset's first-observed delays can plausibly be treated as the customer's actual first delivery delay. Longer panels would also enable the 60-day repurchase window used in the H4 robustness check to be extended without sacrificing sample size, and they would allow the compositional argument advanced in Section 5.2 to be tested directly, since shifts in the retained customer base toward price-sensitive segments take longer than one year to materialize observably in transaction data.

Second, the dataset contains no information on whether the retailer communicated with customers about delayed orders. Some customers may have received proactive notifications or an apology while others did not. Any variation in this recovery behavior would partly account for the observed differences in retention. Future research integrating service recovery data, such as apology emails, compensation offers, and status update timing, with transaction outcomes would help separate the effect of the delay itself from the effect of the retailer's response to it. The managerial implications around proactive recovery developed in Section 5.5 would benefit directly from such data.

Third, the dataset is drawn from a single retailer in the beauty sector. This single-setting design has implications for how the findings should be generalized. The directional disagreements with prior work documented in Sections 5.1 and 5.2 are not evidence that the present design failed to detect existing effects. They suggest that the regularities established in the fulfillment disruption literature do not generalize uniformly across e-commerce settings. The gradual adjustment dynamic reported in Rao et al. (2011), Amorim et al. (2025), and Harter et al. (2025) does not appear here. Whether it appears or breaks down in a given context is itself an empirical question that prior single-study evidence has not been positioned to answer. Future research could systematically vary the empirical setting along dimensions such as switching cost, product substitutability, and price-comparison transparency. Such work would map where the gradual adjustment dynamic holds and where it gives way to the binary response observed here.

Fourth, the analysis is observational. The within-customer specifications for H1 and H2 hold stable individual characteristics constant. They cannot rule out time-varying influences that change for a customer between the pre-delay and post-delay periods. The cross-sectional comparison underlying H4 is more vulnerable. A natural extension would be to identify a quasi-experimental shock to delivery reliability, such as a carrier change, a warehouse expansion, or a sudden capacity constraint affecting some customers but not others. Difference-in-differences applied to such a shock would yield a cleaner causal estimate. The associations documented in this study are consistent with the proposed mechanisms, and a stronger identification strategy would convert them into causal estimates.

6. Conclusion

This study examined how delivery delays influence customers' subsequent purchasing behavior in e-commerce. The empirical setting was a Swedish online beauty retailer operating in a market with low switching costs and high product substitutability. Transaction-level records over a single calendar year were used to test four behavioral hypotheses, with within-customer panel fixed-effects specifications for spending and timing outcomes, and linear probability models for returns and repurchase probability. The behavioral response to delivery delays in this market does not match the pattern established in prior research on fulfillment disruptions. Customers who continued purchasing after a delay did not adjust their spending, extend their purchase intervals, or return more products. The response appeared instead in retention: a substantial share of delayed customers (24.6%) placed no further order during the observation period, reducing effective revenue per delayed customer by 25.7 percent leaving the retailer silently. The cost of late delivery in this setting is concentrated in this exit rather than incremental adjustment among customers who stay. The contribution is to position market structure as a boundary condition on the gradual-adjustment dynamic that prior literature has treated as general. In e-commerce markets where customers can switch retailers without friction, delivery failures do not produce a continuum of reduced engagement. They produce a binary response: customers either tolerate the failure entirely or exit without contact. For retailers operating in such markets, the relevant cost of a delivery

failure is not the operational cost of the delayed shipment. It is the customer lifetime value placed at risk when a delivery promise is broken.

7. Reference list

Akturk, M. S., Mallipeddi, R. R., & Jia, X. (2022). Estimating impacts of logistics processes on online customer ratings: Consequences of providing technology-enabled order tracking data to customers. Journal of Operations Management, 68(6-7), 775-811. https://doi.org/10.1002/joom.1204

Amorim, P., Eng-Larsson, F., & Rooderkerk, R. P. (2025). Navigating online order fulfillment failures: Impacts on future customer behavior and the role of retailer mitigation. Journal of Retailing, 101(3), 382-408. https://doi.org/10.1016/j.jretai.2025.04.008

Anderson, R. E., & Srinivasan, S. S. (2003). E-satisfaction and e-loyalty: A contingency framework. Psychology & Marketing, 20(2), 123-138. https://doi.org/10.1002/mar.10063

Barker, J. M., & Brau, R. I. (2020). Shipping surcharges and LSQ: Pricing the last mile. International Journal of Physical Distribution & Logistics Management, 50(6), 667-691. https://doi.org/10.1108/IJPDLM-09-2019-0292

Cui, R., Li, M., & Li, Q. (2019). Value of high-quality logistics: Evidence from a clash between SF Express and Alibaba. Management Science, 66(9), 3879-3902. https://doi.org/10.1287/mnsc.2019.3411

Deshpande, V., & Pendem, P. K. (2022). Logistics performance, ratings, and its impact on customer purchasing behavior and sales in e-commerce platforms. Manufacturing & Service Operations Management, 25(3), 827-845. https://doi.org/10.1287/msom.2021.1045

Fisher, M. L., Gallino, S., & Xu, J. J. (2019). The value of rapid delivery in omnichannel retailing. Journal of Marketing Research, 56(5), 732-748. https://doi.org/10.1177/0022243719849940

Forbes, L. P., Kelley, S. W., & Hoffman, K. D. (2005). Typologies of e-commerce retail failures and recovery strategies. Journal of Services Marketing, 19(5), 280-292. https://doi.org/10.1108/08876040510609907

Harter, A., Stich, L., & Spann, M. (2025). The effect of delivery time on repurchase behavior in quick commerce. Journal of Service Research, 28(2), 211-227. https://doi.org/10.1177/10946705241236961

Heim, G. R., & Sinha, K. K. (2001). Operational drivers of customer loyalty in electronic retailing: An empirical analysis of electronic food retailers. Manufacturing & Service Operations Management, 3(3), 264-271. https://www.proquest.com/scholarly-journals/operational-drivers-customer-loyalty-electronic/docview/200660912/se-2?accountid=39039

Holloway, B. B., & Beatty, S. E. (2003). Service failure in online retailing: A recovery opportunity. Journal of Service Research, 6(1), 92-105. https://doi.org/10.1177/1094670503254288

Ipsos/Octopia. (2022). E-commerce, marketplaces, and delivery experience. https://www.ipsos.com/en/ecommerce-marketplaces-delivery-experience [Accessed Feb 21, 2026].

Keaveney, S. M., & Parthasarathy, M. (2001). Customer switching behavior in online services: An exploratory study of the role of selected attitudinal, behavioral, and demographic factors. Journal of the Academy of Marketing Science, 29(4), 374-390. https://doi.org/10.1177/03079450094225

Masuch, S., Landwehr, J. R., Flath, C. M., & Thiesse, F. (2024). The faster, the better? The impact of short delivery times on product returns in online retailing. Journal of Retailing, 100(3), 475-485. https://doi.org/10.1016/j.jretai.2024.07.002

Mattila, A. S. (2004). The impact of service failures on customer loyalty: The moderating role of affective commitment. International Journal of Service Industry Management, 15(2), 134-149. https://doi.org/10.1108/09564230410532475

Nguyen, D. H., de Leeuw, S., & Dullaert, W. E. H. (2018). Consumer behaviour and order fulfilment in online retailing: A systematic review. International Journal of Management Reviews, 20(2), 255-276. https://doi.org/10.1111/ijmr.12129

NielsenIQ. (2024). Amazon Prime: The key to Amazon's dominance. https://nielseniq.com/global/en/insights/analysis/2023/amazon-prime-the-key-to-amazons-dominance/ [Accessed Feb 21, 2026].

Oliver, R. L. (1980). A cognitive model of the antecedents and consequences of satisfaction decisions. Journal of Marketing Research, 17(4), 460-469. https://doi.org/10.2307/3150499

Parasuraman, A., Zeithaml, V. A., & Berry, L. L. (1988). SERVQUAL: A multiple-item scale for measuring consumer perceptions of service quality. Journal of Retailing, 64(1), 12-40.

Petersen, J. A., & Kumar, V. (2009). Are product returns a necessary evil? Antecedents and consequences. Journal of Marketing, 73(3), 35-51. https://doi-org.ez.hhs.se/10.1509/jmkg.73.3.035

Rabinovich, E., & Bailey, J. P. (2004). Physical distribution service quality in internet retailing: Service pricing, transaction attributes, and firm attributes. Journal of Operations Management, 21(6), 651-672. https://onlinelibrary-wiley-com.ez.hhs.se/doi/10.1016/j.jom.2003.11.004

Rao, S., Griffis, S. E., & Goldsby, T. J. (2011). Failure to deliver? Linking online order fulfillment glitches with future purchase behavior. Journal of Operations Management,29 (7–8), 692-703. https://doi.org/10.1016/j.jom.2011.04.001

Rao, S., Rabinovich, E., & Raju, D. (2014). The role of physical distribution services as determinants of product returns in internet retailing. Journal of Operations Management, 32(6), 295-312. https://doi.org/10.1016/j.jom.2014.06.005

Scandinavian MIND. (2024). Boozt executive on how LogTech is set to transform logistics as we know it. https://scandinavianmind.com/boozt-executive-tobias-sjolin-on-logtech/ [Accessed Feb 21, 2026].

Suguna, M., Shah, B., Karthik Raj, S., & Suresh, M. (2021). A study on the influential factors of the last mile delivery projects during Covid-19 era. Operations Management Research, 15, 399-412. https://doi.org/10.1007/s12063-021-00214-y

Tamaddoni, A., Stakhovych, S., & Ewing, M. (2016). Comparing churn prediction techniques and assessing their performance: A contingent perspective. Journal of Service Research, 19(2), 123-141. https://doi.org/10.1177/1094670515616376

8. Appendix

Appendix 1 – AI Disclosure

This thesis was written by the author, and all analyses, interpretations, and conclusions are the author's own. Generative artificial intelligence tools were used, including ChatGPT for language refinement and conceptual questions, and Grammarly for grammar correction. AI tools were also used to ask conceptual questions related to statistics and programming to better understand how to implement the analyses.

No AI tool was used to generate empirical results, conduct statistical analysis, or make research decisions. All data handling, model specification, hypothesis testing, and interpretation of results were performed by the author. The author takes full responsibility for the content of this thesis.

Appendix 2 – Additional descriptive analyses

This section reports four additional descriptive analyses of the delivery-delay sample requested during supervision. Together they characterize the volume of purchasing observed before and after the first delay, the timing of delayed orders across the year, the within-customer change in average order value, and the share of delayed customers for whom no purchase is observed before the delay.

Number of orders before and after the first delay

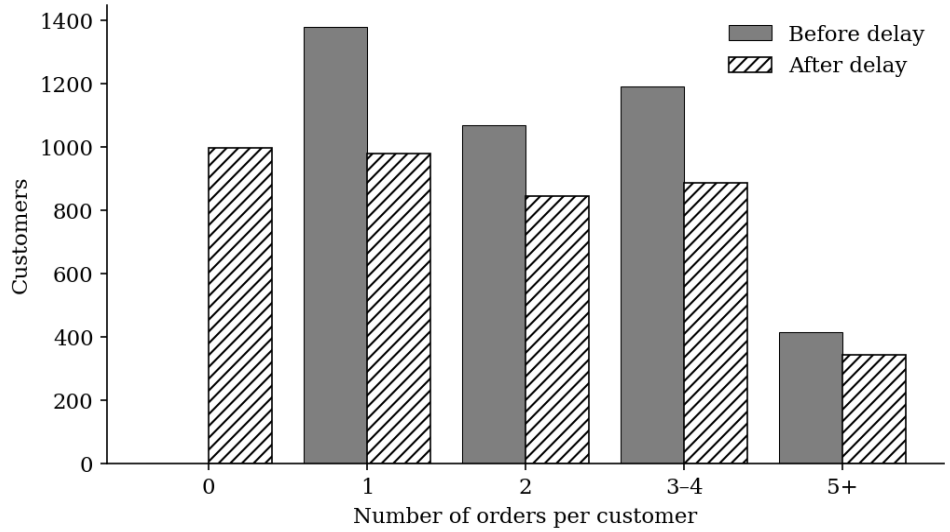
For each of the 4,055 customers who experienced at least one delay, orders are split into a pre-delay period (orders up to and including the first delayed order) and a post-delay period (orders placed afterwards). On average, a customer placed 2.44 orders in the pre-delay period (median 2) and 1.91 in the post-delay period (median 2). The distribution is shown in Table A1 and Figure A1. The pre-delay count is at least one for every customer by construction, because the focal delayed order is assigned to the pre-delay period; the 997 customers (24.6 percent) with zero post-delay orders are the silently exiting customers analyzed under H4.

Table A1. Distribution of orders per delayed customer, before vs. after the first delay

Orders per customer	Before delay	After delay
0	0	997
1	1,381	981
2	1,068	846
3–4	1,192	887
5+	414	344
Mean	2.44	1.91
Median	2	2

Note. N = 4,055 delayed customers. The pre-delay period includes the focal delayed order, so no customer can have zero pre-delay orders.

Figure A1. Number of orders per delayed customer, before and after the first delivery delay.



Timing of delayed orders across the year

Figure A2 shows the number of delayed orders by calendar month. Delays are spread across the year rather than concentrated in a single period, with a visible increase in November and December that is consistent with elevated order volume during the peak shopping season.

Figure A2. Delayed orders by month, 2025

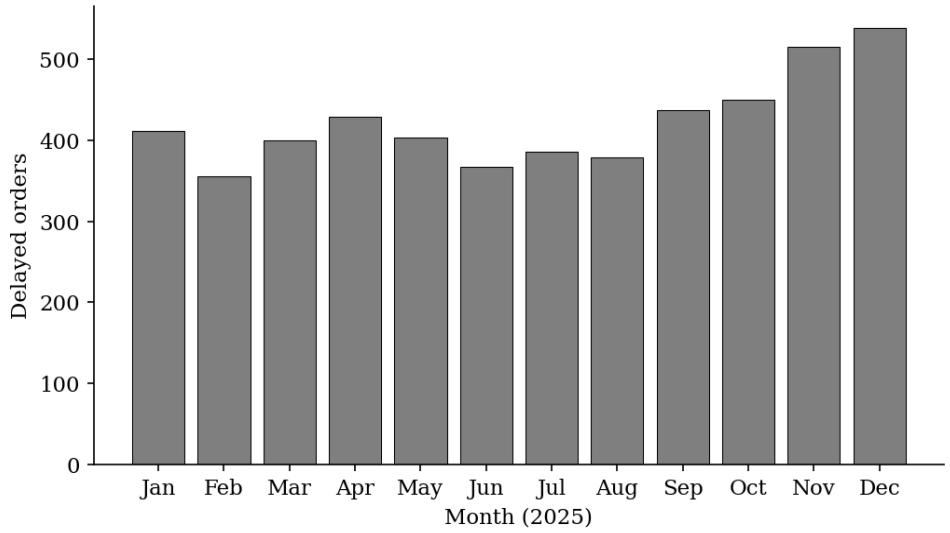


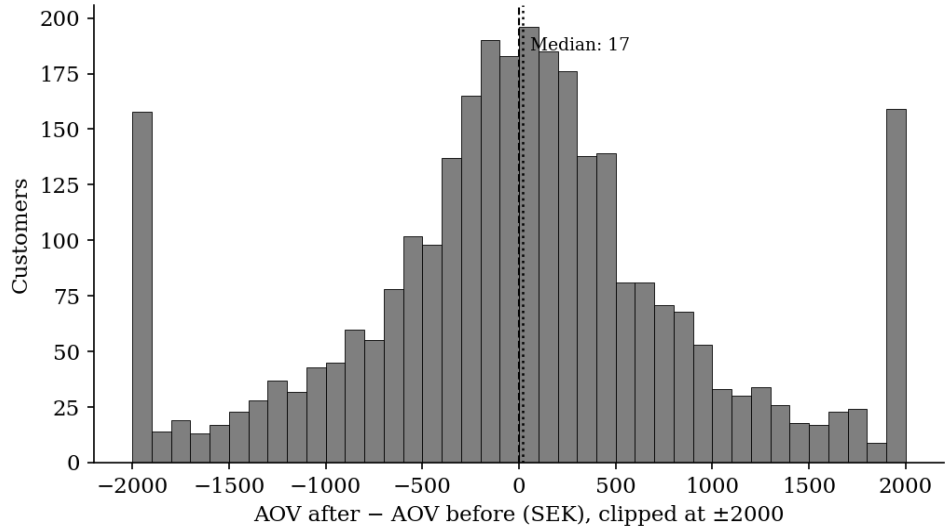
Table A2. Delayed orders by month, 2025

Month	Delayed orders
January	411
February	355
March	400
April	429
May	403
June	367
July	386
August	379
September	437
October	450
November	515
December	539

Within-customer change in average order value

For the 3,058 customers observed in both periods, the change in average order value is computed as post-delay AOV minus pre-delay AOV. The distribution is centred close to zero: the mean change is −12.71 SEK and the median is +16.88 SEK, with 49.0 percent of customers decreasing and 51.0 percent increasing their average order value after the delay. The wide spread (standard deviation 1,590.37 SEK) reflects a small number of large swings in both directions. The near-symmetric distribution around zero is consistent with the null result reported for H1: there is no systematic downward shift in spending among retained customers.

Figure A3. Distribution of the within-customer change in average order value (post-delay minus pre-delay)



Note. N = 3,058 customers with at least one order in both periods. The x-axis is clipped at ±2,000 SEK for readability; the dashed line marks zero and the dotted line the median.

Delayed customers with no prior order

Of the 4,055 delayed customers, 1,381 (34.1 percent) placed no order before their delayed order; that is, the delayed order is the first order observed for these customers in the dataset. Because the within-customer before-and-after design assigns the focal delayed order to the pre-delay period, these customers still contribute one pre-delay observation (the focal order itself) but carry no genuine pre-delay purchasing history against which post-delay behaviour can be compared.

This is handled in two ways. First, the H1 and H2 specifications are estimated only on customers who placed at least one order in both periods, so a customer for whom the delayed order is the first and only pre-delay order enters these models only if a later post-delay order exists. Second, for these customers the pre-delay mean is based on a single order, which adds noise but not systematic bias to the within-customer contrast. The presence of customers without observed prior history also connects to the unobserved-history limitation discussed in Section 5.6: customers who joined before the observation window may have experienced earlier, unrecorded delays, which would attenuate the before-after comparison toward zero.

Table A3. Pre-delay purchasing history among delayed customers

Group	Customers	Share
Delayed order is the first observed order	1,381	34.1%
At least one order before the delayed order	2,674	65.9%
Total delayed customers	4,055	100%

Note. A customer is counted as having no prior order when the date of the first delayed order equals the date of the first order observed for that customer.