Journal of Retailing and Consumer Services 85 (2025) 104315

Contents lists available at ScienceDirect

Journal of Retailing and Consumer Services

journal homepage: www.elsevier.com/locate/jretconser

Striking the right balance: Customising return policy leniency for managing customer online return proclivity and satisfaction

Quang Huy Duong [a,*], Li Zhou [a], Meng Meng [b], Le Thuy An Dang [a], Tiep Duy Nguyen [c]

[a] School of Business, Operations and Strategy, Business Faculty, University of Greenwich, SE10 9LS, London, United Kingdom
[b] School of Management, University of Bath, BA27AY, Bath, United Kingdom
[c] Westcon-Comstor, Tarrytown, NY, 10591, USA

ARTICLE INFO

Keywords:
Return policy
Product return behaviour
Product return intention
Satisfaction
Leniency dilemma
E-commerce

ABSTRACT

E-commerce retailers (e-tailers) commonly adopt generous return policies which not only act as a guarantee to protect the customer's purchase but also help in maintaining their satisfaction. However, this strategy can backfire by encouraging impulsive purchasing behaviour and resulting in a surge of product returns. This creates what is termed the product return policy leniency dilemma. To address that, this paper aims to empirically discover the relationships between product return policy leniency dimensions (time, monetary, effort, scope, and exchange) and two output variables – customer return proclivity and satisfaction. We develop a hybrid method combining machine learning-based data extraction and logistic regression, using a large empirical dataset comprising return policies and reviews from Walmart. The results show that three leniency dimensions – monetary, effort and scope drive customer return proclivity and satisfaction. Time only drives the satisfaction but not proclivity while exchange is in reverse. Our findings imply that customers are amenable to reasonable restrictions in return policies regarding time, effort, and exchange. However, overly lenient return policy terms may fail to adequately address the return policy dilemma. Additionally, partial refund/restocking fees are acceptable for customers with return proclivity if they perceive the initial purchasing cost heavily. Allowing some hazardous/bulky products to be returned under condition may also be seen as a generous term from prospective returners. Overall, e-tailers should display flexibility by incorporating different levels of leniency across five dimensions to balance return satisfaction and intention. This study provides e-retailers a guidance to design an appropriate bespoke return policy.

1. Introduction

Online product returns are common practices in retail. Among the factors driving this phenomenon such as the abundance of product ranges, the diversity of client expectations, and the emergence of e-commerce, return policy (RP) leniency is an important factor (Duong et al., 2022). A RP outlines the terms and conditions for customers to return purchased items. A typical classification framework to decompose RP includes five primary dimensions: time, monetary, effort, scope and exchange (Janakiraman et al., 2016). *Time* refers to the permitted time for return. Longer allowed return periods are regarded as more lenient RPs. *Monetary* relates to the refund amount, expressed as percentage of the purchase price. This dimension relates to the reimbursed value that customers would get back from e-tailers. A full refund is the most leniency RP. *Effort* involves the consumer's effort required by e-tailers to make a successful return, with leniency varying based on the level of inconvenience imposed on customers. *Scope* pertains to the nature of goods that are returnable, where more lenient policies allow for the return of used or hazardous products. Finally, the *exchange* dimension considers whether the e-tailer accepts either exchanges/replacements, grants store credits/gift cards, or cash refunds, with cash refund representing the most lenient RP. This dimension relates to the remedy method that e-tailers offer to a returnable product. While generous RPs enhance customer confidence and satisfaction, they often lead to increased return volumes and operational inefficiencies, creating a *"Return policy leniency dilemma"* for e-tailers.

Many e-commerce retailers (e-tailers) such as Amazon, are offering excessively permissive RPs as an assurance for the lack of pre-purchase

* Corresponding author.
E-mail addresses: quang.duong@gre.ac.uk (Q.H. Duong), li.zhou@gre.ac.uk (L. Zhou), mm3042@bath.ac.uk (M. Meng), lethuyan.dang@gre.ac.uk (L.T.A. Dang), tiepcz@gmail.com (T.D. Nguyen).

https://doi.org/10.1016/j.jretconser.2025.104315
Received 27 February 2025; Received in revised form 27 April 2025; Accepted 29 April 2025
Available online 8 May 2025

physical inspection. High lenient RP is set to enhance customer satisfaction (Heim and Field, 2007) which is defined as a customer's overall evaluation of a business's products and services and mitigate the purchase risk (Rokonuzzaman et al., 2021; Woimant and Steils, 2025). Satisfied customers are more likely to make further purchases as well as to tell others about their experiences via word-of-mouth. However, permissive RPs give space for the rise of returns which its volume might reach above 30 % in some store (Cui et al., 2020; Ketzenberg et al., 2020). Not all e-tailers can afford that influx of returns, especially for small-and-medium enterprises.

E-tailers are striving to discourage returns by imposing some restrictions and even non-returnable policy. For example, retailers like Gap, Old Navy, Banana Republic, and J. Crew have now restricted to from lifetime to one month return period while Zara, Anthropologie, REI, and L.L. Bean impose a fee for mailed returns or allow exchange only (Dickler, 2022; Petro, 2022; Zhang et al., 2023). It is undeniable that stringent RPs would mitigate product return proclivity. Nevertheless, it would harmfully affect genuine customers whose return reasons are legitimate, thus leading to the degradation of customer satisfaction and loyalty (Bower and Maxham, 2012). Since a recent study from Chen et al. (2023) found that legitimate customers perceive higher fairness and have increased purchase intentions under stricter RPs when fraudulent returns are present. Thus, we posit that legitimate customers might still accept some certain RP restrictions. This motivates us to empirically examine how various RP leniency levels affect return proclivity (including return intention and behaviour) and satisfaction among legitimate returners, aiming to provide e-tailers with strategies to navigate this dilemma.

Practitioners and researchers have largely overlooked the perspectives of legitimate customers towards the restricted RPs, focusing primarily on financial benefits and cost management (Abdulla et al., 2019; Loeb, 2022; Mollenkopf et al., 2011). Chen et al. (2023) is one of the few studies to highlight the importance of understanding legitimate returners' views on restricted RPs. Our study, however, differs in two significant ways: (1) We examine different dependent variables – return proclivity and satisfaction versus their perceived fairness and purchased intention and (2) their survey design assumes customers are aware of other's fraudulent behaviours which could be unrealistic. Customers may not always understand why restrictions are in place, or even if they do, they tend to prioritise their own benefits over the reasons for the restrictions. We address this limitation by studying customers' perception towards RPs through customer online reviews. Customer online reviews are appropriate due to several factors: (i) Customers usually review RPs before purchasing, hence reviews can reflect their perception towards RP (Chen et al., 2024), (ii) customers often share their opinions spontaneously in online reviews, reflecting their true feelings and experiences without the influence of experimental settings or survey biases (e.g., the presence of return fraud), (iii) lastly but most importantly, customers typically feel ashamed/embarrassment of sharing return reasons, especially return abuse, hence, those who shared are likely having legitimate reasons.

Despite the increasing academic interest in RPs and their role in shaping consumer behaviour (e.g., Abdulla et al., 2022; Bower and Maxham, 2012; Gäthke et al., 2022; Janakiraman et al., 2016; Janakiraman and Ordóñez, 2012; Rao et al., 2018; Zhang et al., 2017), several key limitations remain in the current literature. First, most existing studies examine one or two leniency dimensions – most commonly monetary, time, or effort – in isolation or paired combinations, often neglecting scope and exchange. Only a few studies (e.g., Abdulla et al., 2022; Janakiraman et al., 2016) adopt a holistic view of RP leniency across all five dimensions. However, even these studies typically rely on binary classifications (e.g., lenient vs. strict), which may obscure the non-linear and multidimensional nature of consumer responses to varying degrees of leniency. Second, although prior research has drawn on established theories such as Expectancy Disconfirmation Theory, Signalling Theory, and Justice Theory (Chen et al., 2023; Dailey and Ülkü, 2018; Pei et al., 2014; Rao et al., 2018; Yan and Cao, 2017), these frameworks have rarely been integrated systematically to examine how different RP leniency features shape return proclivity and customer satisfaction. Notably, the moderating role of return proclivity – distinguishing between return intention and return behaviour – remains underexplored. This gap is particularly significant in online retail contexts, where herd behaviour, information asymmetry, and user-generated reviews strongly influence both perceived performance and post-purchase satisfaction. Third, while satisfaction is a central outcome in consumer behaviour theory, its relationship with RP features has received limited direct attention. Most studies focus on intermediate outcomes such as purchase intention, perceived fairness, or trust, with satisfaction either treated peripherally or not disaggregated by leniency dimension (e.g., Chen et al., 2023; Oghazi et al., 2018; Pei et al., 2014; Wood, 2001). By addressing these limitations, our study contributes to a more granular, theory-driven understanding of the RP dilemma.

In this paper, we aim to examine the relationships among five RP leniency dimensions, customers satisfaction and return proclivity. The goal is to seek the optimal RP leniency level that have positive satisfaction while minimising return proclivity, potentially addressing the dilemma. Two research questions are to be answered:

- *RQ1*. Can e-tailers issue RPs with restrictions without damaging customer overall satisfaction?
- *RQ2*. Do RP leniency dimensions have different effects among return intention and behaviour?

To answer these questions, we first extract five RP leniency dimensions as independent variables, return proclivity and satisfaction as dependent variables. This is done by analysing 13,757 textual RPs and 15,991 customer reviews from Walmart marketplace using machine learning techniques. Then, we analyse the impacts of the independent variables on the dependent variables using logistic regression models. This data-driven approach addresses limitations inherent in traditional methods such as surveys and experimental designs, which often fail to capture authentic managerial dynamics due to vulnerabilities including single-respondent bias and retrospective (Yan et al., 2025). By leveraging large, naturally occurring datasets from real-world platforms, we can explore complex phenomena at scale and uncover nuanced insights into consumer behaviour that are challenging or impossible to detect within controlled experimental environments (Duong et al., 2024; Ketzenberg et al., 2020; Van Nguyen et al., 2020).

Our analysis revealed that monetary, effort, and scope dimensions have heterogeneous effects on return proclivity and satisfaction at different leniency levels. Time does not affect return proclivity, and exchange does not affect satisfaction. Our findings suggest that customers are willing to accept certain restrictions in RPs related to time, effort, and exchange. However, extremely lenient RP terms may not effectively resolve the RP dilemma. Furthermore, customers with return intentions and behaviours perceive sub-optimal RP leniency differently at some leniency levels.

Our analytical approach offers several novel contributions beyond prior studies using the five dimensions by Janakiraman et al. (2016) and Abdulla et al. (2022). First, we leverage large-scale, real-world data from Walmart to capture actual customer sentiment and behaviour, rather than relying on hypothetical scenarios. Second, we disaggregate return proclivity into intention and behaviour, providing a more detailed understanding of the return process. Third, we apply a three-level classification of leniency (low-medium-high), allowing us to detect non-linear effects and uncover optimal leniency thresholds. Fourth, we frame the RP dilemma through the lens of Expectancy Disconfirmation Theory, Signalling Theory, and Justice Theory, illustrating how certain restrictions may serve as credible signals and be positively received by legitimate customers.

Practically, our study offers a strategic guide for e-tailers, particularly small and medium-sized enterprises, by leveraging a machine learning-based text mining approach to analyse real-world RPs and customer reviews. This data-driven framework equips e-tailers with actionable insights to design and customise RPs that balance customer satisfaction with return mitigation, optimising resources while addressing the limitations of overly permissive RPs in managing the leniency dilemma.

The remaining paper is presented as follows. In section 2, we discuss related literature in RP leniency domain and identify the research gaps. In section 3, theoretical framework is presented to build the foundation for our study. In section 4, data collection, variable operationalisation and proposed model are introduced. Then in section 5 and 6, main results and discussions are provided. Section 7 presents managerial and theoretical contributions of this study. Finally, section 7 concludes with summary of the findings and suggests future directions.

2. Literature review

In this section, we review related literature and justify for the research gaps. The extant literature of this study is in the intersection of two domains: (1) RP leniency and (2) consumer behaviour. A table of summary can be found in Online Appendix A.

2.1. RP leniency dimensions

Regarding the first domain, RPs are typically characterised by their leniency, emphasising the convenience and ease with which consumers are allowed to make returns. Lenient RPs have a long history, dating back to the money-back guarantee scheme introduced by the founders of the Walter F. Baker Cocoa firm in 1777 (Tonning, 1956). This monetary concept has been evolving in 1900s to reduce customer purchasing risks and boost sales (Davis et al., 1995; Menke, 1969; Patankar and Mitra, 1995). As retailing grows especially in e-commerce, RP leniency tends to be investigated in various dimensions and operating conditions so that the influence of certain leniency choices could be better understood (Abdulla et al., 2019). Many studies focus on one or two dimensions of leniency, with time, monetary, and effort being the most frequently investigated – either individually or in combination.

Time leniency, for instance, refers to the timeframe specified by e-tailers within which returns are accepted, commonly ranging from 15 to 90 days. Longer return windows are generally considered more lenient. This dimension has been moderately studied on its own, with two representative studies – Rao et al. (2018) and Ülkü et al. (2013) – exploring how greater time leniency can enhance firms' profitability.

Compared to time, monetary leniency is more widely examined, often in terms of full versus partial versus no refunds (Bower and Maxham, 2012; Chen et al., 2024; Hjort and Lantz, 2016; Pei et al., 2014; Shang et al., 2017a, 2017b; Wood, 2001). A more financially lenient policy appears to reduce the overall amount of purchase deliberation, increase perceptions of product quality (Wood, 2001), positively influence consumers' perceptions of RP fairness, trust, and purchase intention (Pei et al., 2014). It could be more profitable for firms when dealing with consumer bracketing (Chen et al., 2024) or having substantial number of repeat customers (Hjort and Lantz, 2016). Conversely, Shang et al. (2017b) point out that the value of a full refund policy to consumers may not be as substantial as one might expect. In specific cases, such as dealing with consumer wardrobing practices, a partial refund (i.e., higher restocking fee) could be more profitable for firms (Shang et al., 2017a).

Effort leniency, on the other hand, indicates the level of effort consumers must exert to return products. Some retailers might require original receipts, packaging or form-filling, while others are more hassle-free. Accordingly, policies that minimise consumer effort are seen as more lenient. The extant literature suggests that reduced effort requirements can lead to higher purchase intentions – contingent on trust and the retailer's control (Bonifield et al., 2010), lower negative attitudes towards the retailer (Dailey and Ülkü, 2018), and reduced perceived purchase risk, thereby enhancing store image and patronage intentions (Rokonuzzaman et al., 2021).

Recent research trends, however, show a shift from focusing on a single dimension to examining the effects of two dimensions. A popular approach is paring monetary with time (Xu et al., 2015; Zhang et al., 2017), or with effort (Hsiao and Chen, 2014; Oghazi et al., 2018), depending on the researchers' experimental design, to account for their interactions or comparative effects. Time (i.e., 2 vs. 7 days) and effort (e.g., form-filling vs. hassle free) are also examined as their own moderators in affecting returns decision by Janakiraman and Ordóñez (2012).

While time, monetary, and effort dimensions have received considerable attention, scope and exchange leniency are less frequently studied. Scope leniency concerns the types of items deemed "return-eligible" – for instance, discounted products may be excluded from returns. Policies with broader return eligibility are seen as more lenient. Meanwhile, exchange leniency relates to the form of reimbursement, where policies providing cash refunds are the most lenient, followed by store credit, while exchanges represent the least lenient option. Chen et al. (2023) is among the few that consider scope, alongside monetary and time, under the assumption that some consumers may feel fairer if they are aware of the presence of return fraudsters. Heim and Field (2007), on the other hand, include exchange in addition to monetary, time, and effort, by examining whether firms should offer full cash or store credit only. Notably, Abdulla et al. (2022) and Janakiraman et al. (2016) stand out as the only studies cover all five dimensions, exploring heterogeneities in customers' behaviours.

To ensure a comprehensive analysis, this study adopts an all-encompassing approach by evaluating five dimensions of RP leniency. In contrast to prior research that often relies solely on literature or experimental designs, we measure these dimensions using actual RP texts written by e-tailers. Moreover, while many studies classify leniency using binary indicators (e.g., Abdulla et al., 2022; Janakiraman et al., 2016), this approach can overlook potential non-linear effects on return proclivity and satisfaction.

To capture more nuanced patterns, we categorise each RP dimension into three levels of leniency: low, medium, and high. These levels, presented as low–medium–high for generalisation, are grounded in findings from previous studies and shaped by the distribution of RP features observed in Walmart marketplace data (see Table 6). This dual approach ensures the categorisation is both theoretically informed and reflective of real-world practice. This classification is well-suited for identifying the optimal balance between return proclivity and satisfaction. A comparable method was effectively applied by Ramanathan (2011), who incorporated a medium-risk level into Finch (2007) customer risk perception model.

2.2. Consumer product returns research

In the second domain, consumer product return can be classified into two segments: (1) return intention and behaviour and (2) cognitive and affective responses (Abdulla et al., 2019). The first segment relates to how they form return proclivity from their perception on RP. Product return proclivity refers to a customer's likelihood of returning a purchased item. (Janakiraman et al. (2016) and Janakiraman and Ordóñez (2012) are among the few who examine the impact of RP leniency on return proclivity, while (Bonifield et al., 2010; Chen et al., 2024; Gäthke et al., 2022; Hsiao and Chen, 2014; Wood, 2001) study return actions and decisions. Our research, however, examine both return intention and return behaviour separately under the umbrella term *"return proclivity"*. This distinction highlights the potential discrepancy between consumers' stated intentions and their actual behaviours (Arts et al., 2011; de Mesquita et al., 2023). As we utilise online reviews, where information about others' decisions is readily available, there is also a high chance that the consumers would form their behaviours based on others' decisions (Rejikumar et al., 2022). This tendency, known as "herd behaviour", explains individuals' preference to follow the decisions of others, even when they have their own private information to consider. Customers can rely on existing reviews to make purchase and return decisions (Minnema et al., 2016), which potentially widens the gap between their stated intentions and actual actions.

Meanwhile, the second segment covers customers' cognitive responses about the products and services acquired through information and their corresponding affective responses. Within this domain, perceived product and service quality are the most common target variable in the literature (Abdulla et al., 2022; Wood, 2001; Xu et al., 2015; Zhang et al., 2017). The relationships between RPs and customers' perceptions of risk, fairness and trust have also been extensively studied (Bonifield et al., 2010; Bower and Maxham, 2012; Chen et al., 2023; Oghazi et al., 2018; Pei et al., 2014; Rokonuzzaman et al., 2021).

Despite the direct impact of customer satisfaction on repurchase and loyalty (Fornell, 1992), the impact of RP on satisfaction has received relatively little attention. Customer satisfaction is defined as a consumer's overall evaluation of a company's product and service offerings. Heim and Field (2007) have examined the link between RPs and customer satisfaction using a sample of over 1000 e-service operations, while Dailey and Ülkü (2018) examined whether customers emerge negative attitude when their returns are in denial. However, while Heim and Field (2007) only considered RP as a small aspect of their main research about e-service quality, both studies did not study all leniency dimensions in a granular way. Building on this foundation, we aim to not only examine the leniency level of different dimensions on customer satisfaction but also the moderating role of return proclivity in this relationship which will be further elaborated in the next section.

3. Theoretical framework

The well-established Expectancy Disconfirmation Theory (EDT) provides a robust framework for understanding consumer behaviour, particularly in the context of e-tailers' RPs (Anderson, 1973). The theory posits that consumers' satisfaction is shaped by the comparison between their initial expectations and the actual performance they experience (Oliver, 1980). The core constructs of EDT include: (1) ***Expectations***, which are the beliefs or anticipations a consumer has before encountering a product, service, or experience; (2) ***Perceived performance***, referring to the consumer's perception of the actual performance of the product, service, or experience; (3) ***Disconfirmation of beliefs***, which involves the judgments or evaluations made by comparing the actual performance to the initial expectations; and (4) ***Satisfaction***, which is the ultimate outcome of this comparison, reflecting the consumer's contentment with the product, service, or experience after direct experience with it.

In the context of this study, consumers form their expectations based on factors such as time, monetary cost, effort, scope, and exchange leniency in the RPs. These expectations not only shape the return process but also influence the broader perception of the e-tailer's overall performance (Dailey and Ülkü, 2018). Meanwhile, return proclivity, which includes both return intention (the consumer's plan to return a product) and return behaviour (the actual act of returning), is closely tied to perceived performance.

Perceived performance refers to how well the return process meets or falls short of the consumer's expectations. In this case, return intention can be viewed as a form of perceived performance as it reflects an anticipatory judgment based on consumers' expectations of how the return process will unfold. It influences their willingness to engage in return behaviour, which, in turn, affects their satisfaction with the product, service, and overall experience. For consumers who move beyond intention to indicate actual return behaviour, their perceived performance becomes more tangible. This direct experience with the RP shapes their satisfaction not only with the policy but also with the e-tailer. Thus, both forms of return proclivity are interrelated, with the alignment – or lack thereof – of the return process with consumer expectations influencing satisfaction.

As mentioned, disconfirmation arises when there is a divergence between expectations and reality, leading to either positive (e.g., a return process outperforms expectations) or negative (e.g., a return process falls short of expectations) disconfirmation. These types of disconfirmations subsequently affect consumer satisfaction with both the RP and the e-tailer. On one hand, a lenient policy, offering easy returns and minimal restrictions, could reduce perceived purchase risk and foster a sense of security among consumers. In e-commerce, information asymmetry between consumers and e-tailers is a common issue since customers cannot fully inspect products before purchasing and instead rely on descriptions, images, and reviews. It is also highlighted that this information gap leads to heightened consumer uncertainty, which can negatively impact satisfaction (Yan and Cao, 2017). In response to such uncertainty, lenient RPs, according to signalling theory, can serve as signals of product quality, trustworthiness, and customer service (Rao et al., 2018). These policies indicate that the e-tailer is committed to resolving any issues, creating an environment where consumers feel secure and are more likely to perceive the e-tailer as reliable. This fosters positive disconfirmation. When the return process meets or exceeds expectations, it generates satisfaction and reinforces trust in the e-tailer. In contrast, restrictive policies may signal a lack of confidence in product quality or an emphasis on cost-saving rather than consumer-centric values, potentially leading to negative disconfirmation.

Justice theory, however, offers another important lens for understanding consumer satisfaction with RPs. This theory focuses on how consumers evaluate fairness based on distributive, procedural, and interactional justice, which refers not only to the perceived fairness of outcomes but also to the fairness of the processes leading to outcomes (including the clarity and accessibility of the RP) and the fairness of interpersonal treatment during the return process (Greenberg, 1990). In this sense, while lenient RPs are often perceived as fair, offering customers flexibility with minimal hassle, they can also raise concerns about fairness if consumers feel the policy is too permissive. In particular, if the policy is excessively lenient, consumers might perceive it as an invitation for abuse, leading to negative perceptions about fairness. On the other hand, restrictive policies can be perceived as fair if consumers view them as justified, particularly when the policies are clearly communicated and applied uniformly (Chen et al., 2023). Furthermore, justice theory suggests that fairness perceptions play a significant role in satisfaction (Pei et al., 2014). Consumers who perceive a RP as fair are more likely to experience positive disconfirmation, even if the process has some level of restrictions. Conversely, unfair policies can lead to negative disconfirmation, reducing satisfaction with both the policy and the e-tailer.

In sum, RPs should strike a balance between leniency and restriction. While lenient policies can reduce perceived risk and foster positive disconfirmation (i.e., higher satisfaction and higher return proclivity), overly lenient policies may raise concerns about fairness. Conversely, restrictive policies may enhance fairness perceptions if they are perceived as justified, but they risk creating negative disconfirmation (i.e., lower satisfaction) if seen as overly punitive. These connections provide a basis for investigating the research gap regarding the level of RP restriction/leniency can impose to reduce return proclivity while maintaining customer satisfaction. Drawing from our theoretical foundation, the relationships between RP leniency dimensions, return proclivity, and satisfaction are outlined in the theoretical framework in Fig. 1.

4. Research methodology

4.1. Data collection

We developed a crawler using Scrapy Python package to scrape all possible RPs from Walmart marketplace in 2022. At the time of data

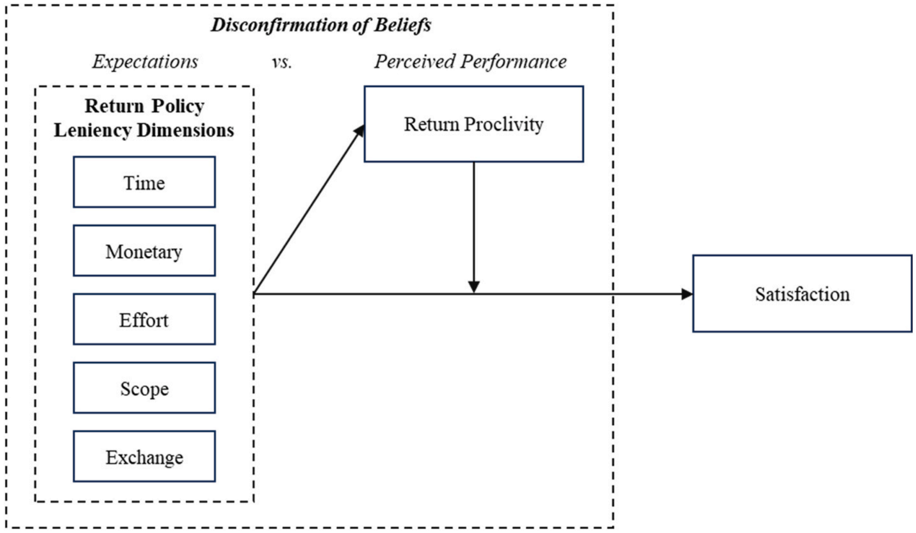

Fig. 1. Theoretical framework.

collection, Walmart publicly displayed the RPs of e-tailers when such information was not readily available on other marketplace platforms like Amazon or eBay. While Amazon provides more standardised RPs across e-tailers, limiting variations in leniency levels, eBay displays RPs at the product level rather than at the retailer level, which is not the focus of our research. Walmart's textual datasets provide a unique opportunity to extract the RP leniency levels across five different dimensions in one place. Additionally, customers can easily browse and scrutinise RP signals, which may influence their return behaviour and overall satisfaction. This study assumes that customers are aware of RPs at the time of purchase, consistent with literature suggesting that RPs function as pre-purchase signals of quality and risk reduction (e.g., Chen et al., 2024; Pei et al., 2014; Robertson et al., 2020; Xu et al., 2015).

After removing many invalid RPs such as "N/A", "TEST SELLER", "Test Channel", 13,757 e-tailers are valid for further analysis. Within each e-tailer, we also collect their reviews to measure customer reaction towards RP leniency. A total of 25,991 reviews were collected after removing some duplications. These reviews are from 3172 out of 13,757 e-tailers since not all retailers have received customer reviews. The collected text data is cleaned using the process in Online Appendix B. Furthermore, we also collect data on the primary product categories each e-tailer specialises in. In total, we have identified 20 major categories, as defined by Walmart (see Online Appendix E).

4.2. Variables identification and measurement

Table 1 summarises the values to be used and the methods for generating them for the dependent, independent, and control variables examined in this study. The description and justification for each variable will be outlined in the following sections.

4.2.1. Dependent variables

- **Customer return proclivity**

The dependent variable, customer return proclivity, is captured from customer reviews using information retrieval techniques. It comprises two main human-machine collaboration steps of (1) keyword matching and (2) human annotation. This two-step approach, ensuring the accuracy while minimises manual effort, has been widely adopted by operations and marketing research (Duong et al., 2024; Ko and Bowman, 2023; Tirunillai and Tellis, 2014).

We first compile a comprehensive list of return-related keywords based on Walmart RP website and our research team's expertise (Fig. 2).

Table 1
Summary of variables measurement in this study.

Type	Variable	Values	Method
Dependent	Customer return proclivity	Contains three classes – Class 0 (No return intention), Class 1 (Return intention) and Class 2 (Return behaviour)	Keyword matching and human annotation
	Customer satisfaction	Contains two classes – Positive and Negative	Pre-trained BERT-based sentiment analysis, polarity scaling and machine annotation
Independent	Time leniency level Monetary leniency level Effort leniency level Scope leniency level Exchange leniency level	Contains three levels – Level 1 (Low), Level 2 (Medium) and Level 3 (High)	BERTopic modelling and human annotation
Control	Product category	Contains 7 categories – (1) furniture and furnishing; (2) recreation, sport and culture; (3) grocery and food; (4) heath, personal care and beauty; (5) electronics; (6) vehicles; and (7) fashion	Extracting directly from the dataset and classifying using COICOP 2018 Category
	RP length	Continuous number of words	Counting method
	RP sentiment	Continuous number ranging from −1 (very negative) to 1 (very positive)	Pre-trained BERT-based sentiment analysis, polarity scaling
	RP subjectivity	Continuous number ranging from −1 (very subjective) to 1 (very objective)	TextBlob
	Review length	Continuous number of words	Counting method
	Review seasonal effect	Contains 12 dummy variables from January to December	Extracting directly from the dataset

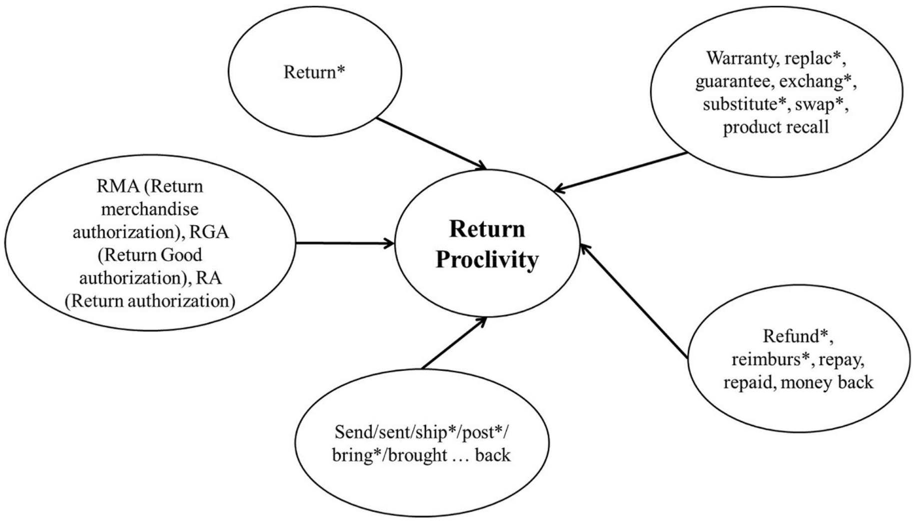

Fig. 2. Keywords indicating customer return intention (the asterisk * shows various word forms).

If a review contains any of these keywords, we denote as returns proclivity. Otherwise, we denote as no return intention. As a result, there are 3829 (14.73 %) reviews denoted as return proclivity and 22,162 (85.27 %) reviews denoted as Class 0 – no return intention (i.e., non-returners). This step effectively reduces the manual workload from 25,991 reviews to 3829 reviews. Machines excel at this task because they follow clear and standardised rules to detect predefined return-related terms, ensuring consistency.

Second, in line with our research questions, we further classify 3829 reviews with return proclivity into two classes (see Table 2 for examples):

(1) Class 1 – return intention (i.e., prospective returners) are any reviews which do not indicate a confirmation of past/ongoing product returns.
(2) Class 2 – return behaviour (i.e., returners) are any reviews which indicate a confirmation of past/ongoing product returns.

This task is challenging as customers may diversify their used language in expressing return intention or behaviour. For example, a review may state, *"I am returning the product"* in one sentence but later contradict it with *"But I changed my mind and kept it."* Such nuances necessitate human interpretation rather than automated classification. This requires a carefully reading and analysing each return-related review.

First, to reduce human fatigue in reading long texts, we filtered and kept only the sentences contained return-related keywords in each review utilising the comprehensive list in Fig. 2. Second, to reduce human bias, we recruited two independent annotators for this task. Human annotators can classify the filtered reviews into return intention or return behaviour, as this distinction often requires contextual understanding beyond what automation can accurately capture. Two annotators achieved a high Cohen's Kappa score of 0.901 indicating perfect inter-annotator agreement (Cohen, 1960). Finally, we cross-checked and discussed assigning the final labels for the mismatched ones. Among 3829 (14.73 %) reviews with return proclivity, we found that 2426 reviews indicate return intention (Class 1), and 1403 reviews indicate return behaviour (Class 2).

Table 2
Examples of reviews in class 1 and 2.

Class	Review contents
1	I *have not returned* the item yet. I will need to *consider if the return* is easy to complete or if some policy, return fee, etc.
1	I *want my money back* and I *will return* this very small hand towel.
2	So, I *returned it* and paid $3.50 for send it back.
2	*The return went well*. Already got an email from Walmart about a future credit.

- **Customer satisfaction**

Customer satisfaction is characterised as a customer's total assessment of an overall product and service offerings by the businesses. Thanks to the development of the Internet, retailers could gain more insights into how satisfied their customers are via feedback from the textual reviews. Commonly, customer satisfaction is captured by the rating given by customers when they write a review. From large online marketplace like Walmart, eBay or Amazon, the range can be from 1 star with low dissatisfaction to 5 stars with high satisfaction. However, it was evident that star rating may not always align with review text, and they are ambiguous to multiple interpretations such as 3-star reviews (Mudambi et al., 2014). Instead, we use sentiment as a proxy to measure customer satisfaction, evident from previous studies (Chandrashekaran et al., 2007; Chen et al., 2020; Wang et al., 2021).

We propose a novel approach to measure their sentiment through their review text. A review text can be used to detect customers' sentiment where they satisfied or dissatisfied with product and service aspects. Customer overall satisfaction depends on the aggregated sentiment across various aspects (Qiu et al., 2018). Hence, review text is a potential and rich source for accurate satisfaction measurement. We apply a RoBERTa pre-trained machine learning-based sentiment analysis[1] to calculate sentiment scores (Barbieri et al., 2020). This model was proved to achieve a high accuracy score of ~73 % which is significantly higher than some standard models such as SVM or FastText on sentiment analysis tasks. The model is selected as it utilises the breakthrough Bidirectional Encoder Representations Transformers (BERT) language model which takes into consideration the contextual environment to understand different meanings of a word (Devlin et al., 2019). For example, the word *"interesting"* may have both positive and negative elements in different contexts.

First, we employ the pretrained model, which generates three sentiment scores for each review: positive, negative, and neutral. As neutral sentiment does not provide substantial managerial insights, we exclude it from further analysis. Second, to transform the remaining two sentiment scores into a single satisfaction variable from −1 (negative) to

[1] Available at Huggingface Python package: cardiffnlp/twitter-roberta-base-sentiment · Hugging Face.

Table 3
Examples of positive and negative reviews.

Reviews	Positive score	Negative score	Overall satisfaction
Customer service was good, I returned product to store and got an email that I will be seeing my refund soon.	0.9276	0.0061	0.9868
The item was received quickly. When I returned the item, I received a refund immediately.	0.8097	0.0056	0.9862
The necklace was of poor quality and looked cheap, was not worth the money so I returned it to the seller. I never received a refund. Emailed the seller and did not receive a reply.	0.0046	0.9383	−0.9902
I returned the product, and they refused to provide a free shipping label to return it, nor would they pay me back to return their product that I was not happy with. I would never buy from them again. I paid $29 to have it shipped back.	0.0052	0.9299	−0.9887

Table 4
Lists of RP leniency keywords.

Dimension	Keywords
Time	Hour*, day*, date*, week*, month*, year*
Monetary	Refund*, money, fee, monetary, restocking fee, shipping charge, shipping fee, handling charge, handling fee, cost
Effort	Return merchandise authorization (RMA), return authorization (RA), hassle, trouble, effort, receipt, tag*, packag*, question, form, fill, cancel, shipping return, procedure
Scope	Customiz*, accept*, eligibl*, custom*, (dis)approv*, satisfy*, (dis) allow*, discount*, clearance, sales, promotion*, refus*, reject*, declin*, special order, personaliz*, made to order, consumable, one-of-the-kind, (un)used, condition*, unworn, defect*, (un)open*, any product
Exchange	Exchang*, replc*, voucher, coupon, gift*, cash, warrant*, repair*, guarantee*, credit card, credit*, point

The asterisk (*) sign is used to capture all possible family words of the designated keyword (e.g., eligib* can catch both eligible and eligibility).

1 (positive) for logistic regression later, we apply a scaling formula as suggested by Siering et al. (2018) below:

$$\mathit{Overall\ Satisfaction} = \frac{\mathit{Positive} - \mathit{Negative}}{\mathit{Positive} + \mathit{Negative}}$$

As a result, we extract 16,895 positive reviews (72.81 %) and 9096 negative reviews (27.19 %). The examples of positive, negative and overall satisfaction scores are shown in Table 3.

4.2.2. Independent variables – RP leniency dimensions

- **BERTopic description, justification and preparation**

In this section, we measure the leniency level of five dimensions: time, monetary, effort, scope and exchange following the unified classification framework from Janakiraman et al. (2016). These dimensions represent core, retailer-controlled elements of RPs and reflect key trade-offs between customer convenience and operational cost. Other factors, such as return initiation ease or customer service quality, are considered part of broader service delivery and are indirectly captured through customer sentiment in reviews. As observed, each e-tailer compose their RP which captures RP leniency dimensions in an unstructured and inconsistent way. Hence, topic modelling is appropriately applied to extract the common patterns in each leniency dimension. Topic modelling is an unsupervised machine learning technique used to statistically discover dense semantic topics (i.e., RP leniency terms) and clusters a large number of documents (i.e., a corpus) into these topics. Traditionally, probabilistic topic models such as Latent Semantic Indexing or Latent Dirichlet Allocation have gained momentum to derive prominent topics from a corpus. These models are widely applied since their mechanism utilises the vocabulary list and its occurrence in a whole corpus, which is straightforward and interpretable. However, they fall short as an effective topic extraction tool in distinguishing a word in different context. For example, the meaning of the word *"return"* may vary among giving back (e.g., product returns), financial return (e.g., earnings), or coming back (e.g., return to work). On another note, they are expensive in terms of computational resources and time due to a sparse matrix they create from short-to-medium textual RPs.

To address these drawbacks, we apply BERTopic which utilises the infrastructure of the novel BERT model to extract the prominent topics for RPs (Grootendorst, 2022). We use the default parameters for BERTopic models. The model's initiative is the ability to embed contextual information into the representation vector of a word/sentence/document by scanning bidirectionally (i.e., left, and right direction). A word/sentence/document is quantified into minimum 384-dimensional dense vector representations which overcome the shortcomings of a sparse matrix. Furthermore, the model does not require to pre-define numbers of topics which is suitable for exploring all possible expressions of RP leniency. A process of BERTopic consists of four steps which are (1) document embedding, (2) dimensionality reduction, (3) clustering, and (4) topic representation generation. A detailed explanation of each step can be found in Online Appendix C.

Any information which does not cover the leniency dimensions is considered irrelevant. For example, some sentences such as *"at …, we help you choose fine diamonds and jewelry that you have set your heart on and make sure your expectations are met or exceeded"* or *"we take great pride in our reputation for excellent quality merchandise and value"* are generic and may contaminate the results. To ensure the reliability of the results, we compile five lists of keywords for the five dimensions based on the literature, the authors' expertise, and initial inspection from the dataset (Table 4).

- **BERTopic implementation**

We apply BERTopic to all 13,757 RPs although only 3172 e-tailers have customer reviews. Because topic model requires a large dataset to discover all possible latent topics, omitting any sample may lead to overlooking key leniency information. Since BERTopic tends to work more effective with shorter and dense documents rather than long documents, we matched each keyword to each RP. Then, we respectively filtered and retained words surrounding the matched keywords with a window sliding of 10 words[2] (i.e., 5 words to the left and 5 words to the right). If a dimension is not specified in a RP, the cell will be left blank. We produced five separated corpus representing time, monetary, effort, scope, and exchange. BERTopic was fit into each newly born corpus separately, and we discovered 45 topics in time dimension, 34 topics in monetary dimension, 41 topics in effort dimension, 56 topics in scope dimension and 35 topics in exchange dimension (see Online Appendix D). The quality of BERTopic models is evaluated using the common coherence score, C_v, which is highly interpretable and aligns

[2] A sliding window of 10 words is a common practice in natural language processing techniques such as LDA, Skip-gram, GloVe or CboW for short texts (Bicalho et al., 2017). We attempted to vary this sliding window, but they did not yield better results. This setting is appropriate to avoid overlapped topics across five dimensions.

Table 5
Coherence scores (C_v) from BERTopic modelling.

	Time leniency	Monetary leniency	Effort leniency	Scope leniency	Monetary leniency
Coherence score (C_v)	0.5024	0.6485	0.5148	0.5601	0.5458

Table 6
RP leniency operationalisation and frequency.

Dimension	Operationalisation	Full set frequency	Subset frequency	T-test[a]
Time leniency	Level 1: Up to 14/15 days (T1)	1240 (9.01 %)	297 (9.36 %)	*0.62*
	Level 2: Up to 30 days (T2)	9777 (71.06 %)	2246 (70.80 %)	*−0.29*
	Level 3: More than 30 days (T3)	2740 (19.92 %)	629 (19.82 %)	*−0.11*
Monetary leniency	Level 1: Shipping & Restocking fee/Partial refund (M1)	4579 (33.28 %)	1038 (32.72 %)	*−0.60*
	Level 2: Shipping/Restocking fee (M2)	3394 (24.67 %)	731 (23.04 %)	*−1.92*
	Level 3: Full Refund (M3)	5784 (42.04 %)	1403 (44.23 %)	*2.25*
Effort leniency	Level 1: Conditional (e.g., original tag, package or receipts) (Ef1)	5361 (38.96 %)	1250 (39.40 %)	*0.46*
	Level 2: Only need to fill forms or contact sellers (Ef2)	3134 (22.78 %)	726 (22.88 %)	*0.13*
	Level 3: Hassle free (Ef3)	5262 (38.24 %)	1196 (37.70 %)	*−0.57*
Scope leniency	Level 1: Some products cannot be returned (S1)	4914 (35.71 %)	1087 (34.26 %)	*−1.54*
	Level 2: Some products can be returned under certain conditions (e.g., unopened, unwashed, unused) (S2)	6101 (44.34 %)	1429 (45.05 %)	*0.72*
	Level 3: Any product can be returned (S3)	2742 (19.93 %)	656 (20.68 %)	*0.95*
Exchange leniency	Level 1: Exchange/Replace/Repair (Ex1)	4430 (32.20 %)	1094 (34.48 %)	*2.48*
	Level 2: Gift card/Store credit (Ex2)	223 (1.62 %)	53 (1.67 %)	*0.20*
	Level 3: Cash refund (Ex3)	9104 (66.17 %)	2025 (63.83 %)	*−2.50*

[a] Numbers in italic indicates no statistical significance at level 0.1.

well with human judgment (Röder et al., 2015). Table 5 presents strong coherence scores ($C_v \geq 0.5$) across five leniency dimensions, as suggested by An et al. (2023) and Chatterjee et al. (2024), indicating that the model has extracted high-quality and meaningful results.

Next, each extracted topic will be assigned with a leniency level. From extant literature, RP leniency is usually measured by a binary variable of low and high degree (Janakiraman et al., 2016; Janakiraman and Ordóñez, 2012; Shang et al., 2017b). However, evident from the extracted topics, we find that RP leniency can be classified into three levels of (1) low – (2) medium – (3) high (Table 6). This classification is appropriate to explore the optimal level for balancing return proclivity and satisfaction. A similar approach has been successfully implemented by Ramanathan (2011) in adding medium risk level into customer risk perception model of Finch (2007). We label and apply a rigorous evaluation process to validate the accuracy by inviting two experts to check the label independently using information in Table 6 and the BERTopic results in Online Appendix D. After that, we brainstorm and cross-check to come up with the final labels.[3]

This two-step approach – BERTopic modelling and expert labelling – ensures both efficiency and accuracy in coding the leniency dimensions. BERTopic is employed to cluster RPs into thematically similar groups, providing structure to the unstructured text data. However, as topic modelling techniques do not inherently assign leniency levels, independent experts subsequently cross-check and label these clusters based on predefined theoretical constructs. This hybrid approach (machine-assisted topic modelling followed by expert labelling) is a necessary approach when using topic modelling (Desiraju et al., 2024; Duong et al., 2024; Tirunillai and Tellis, 2014).

Finally, we match the labelled topics to customer reviews based on e-tailer's names. Note that 25,991 reviews are from 3172 e-tailers only. We are aware that the subset of RPs may not statistically represent the whole dataset as it is not randomly selected. Therefore, we first inspect the distributions of the subset and the full set and observe consistent distributions across five dimensions (see Table 6). To further confirm with higher level of confidence, we apply T-test to examine whether the sample has an equal mean with the population. Table 6 shows that all dimensions have homogeneous means with the full set of RPs (p-value $\geq$0.01). The sample of 3172 e-tailers is integrated with the review-level dataset for further analysis.

4.2.3. *Control variables*

We control our model with several variables that might have the influences on the customer return proclivity and satisfaction: Product category, RP length, RP sentiment, RP subjectivity, review length, and seasonal effect (Cui et al., 2020; Heim and Field, 2007; Minnema et al., 2016; Xu and Lee, 2020).

According to Walmart, there are 20 major departments for product categories included in our dataset. However, to reduce the skewness and the risk of multi-collinearity among categories, we use the COICOP2018 guideline from United Nation Statistic Division to further group them into 7 main categories: (1) furniture and furnishing; (2) recreation, sport and culture; (3) grocery and food; (4) heath, personal care and beauty; (5) electronics; (6) vehicles; and (7) fashion (see Online Appendix E).

RP length is measured by the number of words written in a RP. *RP sentiment* is ranged from 1 (very positive) to −1 (very negative) using the same BERT-based sentiment and polarity scaling method in section 4.2.1. *RP subjectivity* refers to a value that indicates the degree to which the text expresses personal perspectives. This variable is measured using the common Python package – TextBlob, ranging from 0 (very objective) to 1 (very subjective). These three variables represent the linguistic characteristics of each RP. Studies found that the linguistic styles affect customer overall satisfaction and behaviours (Xu and Lee, 2020; Zhao et al., 2019).

Review length is measured by the number of words written in a review. Mudambi and Schuff (2010) suggested that the usefulness and richness of a review can be evaluated through review length. Extended

[3] Noted that some e-tailers do not specify all five dimensions in their RPs. These missing values can be classified based on Walmart standard RP for marketplace items (Time – Up to 30 days; Monetary - Shipping/Restocking fee; Effort – Hassle free; Scope – Some products cannot be returned such as rechargeable battery, household chemicals; Exchange – Cash refund. More details can be found at https://corporate.walmart.com/policies#return-policy).

reviews provide customers with greater opportunity to articulate their emotions and intentions to return. This is particularly evident among returning customers, whose emotional states may lead them to express themselves more intensely to alleviate cognitive dissonance and suppress intrusive thoughts (Duong et al., 2024; Sahoo et al., 2018).

Review seasonal effect is measured by 12 dummy variables of the posted month of the review (January to December). Literature demonstrated that season effects have a direct impact on returns volume (Cui et al., 2020). For example, during festival seasons such as Christmas, customers may post more return-related reviews as they are rush in getting return/exchange.

Past literature has identified seller reputation as a potential confounding variable that may influence both return proclivity and customer satisfaction (Pei et al., 2014; Robertson et al., 2020; Walsh et al., 2016). However, in our study, seller reputation, measured through star ratings, may not pose significant endogeneity concerns. This is because major marketplaces such as Walmart tend to remove low reputation e-tailers through their seller performance standards (Walmart, 2025), which leads to a relatively homogenous sample of sellers with moderate to high reputations. This reduces the confounding effect of reputation on return proclivity and satisfaction. Given these considerations, the influence of seller reputation in our dataset is likely to be less pronounced compared to other contexts where a broader range of seller quality is present.

4.3. Descriptive analysis

In this section, we integrate the subset of five independent variables of *RP leniency dimensions*, RP-related control variables – *RP length*, *RP sentiment*, and *RP subjectivity* into the review-level dataset containing two control variables – *Review length* and *review seasonal effect* and two output variables – *return proclivity* and *satisfaction*.

As we intend to use logistic regression models, it is crucial to verify the underlying assumptions, specifically the absence of multicollinearity and strongly influential outliers. To ensure these conditions are met, we first perform correlation analysis, Variance Inflation Factor (VIF) for checking multicollinearity. A variable is not collinear with other variables if they have low correlation coefficients (<0.5) with other variables and VIF score ≤10. All included variables satisfied these conditions, confirming no significant multicollinearity concerns in our dataset. A table of descriptive summary, VIF and correlation can be seen in Appendix I.

Second, strongly influential outliers can be detected using Cook's distance and standardised residuals. Any sample simultaneously has Cook's Distance >4/N, where N is the total number of samples, and absolute standardised residual values > 3 is deemed as strongly influential outliers. We detected 3.67 % samples (threshold of 5 %) in our dataset which can be removed without losing the generalisation of the model (Van Nguyen et al., 2020).

4.4. Model development

Multinominal logistic regression model for *return proclivity* and binary logistic regression model for *satisfaction* are employed to study the multivariate relationships and answer three research questions. Model (1) can be formulated as:

$$Proclivity_i = \beta_0 + \beta_1\, RPL_i + \beta_2\, PC_i + \beta_3\, RPLS_i + \beta_4\, RL_i + \beta_4\, SE_i + \varepsilon_i \tag{1}$$

where $Proclivity_i$ indicates whether a review by user i is no return intention (non-returners), return intention (prospective returners) or actual return behaviour (returners). RPL_i is a vector of dummy variables for all 5 return leniency dimensions (3 dummy variables per dimension). PC_i is vector of dummy variables for 7 main product categories. $RPLS_i$ is a vector of 3 numerical RP-related control variables, which are RP length, RP sentiment, and RP subjectivity. RL_i is a numerical review length. Finally, SE_i is a vector of dummy variables for 12 months from January to December.

To examine customer satisfaction, we turn gears to formulate binary logistic regressions models as below:

$$Satisf_i = \beta_0 + \beta_1\, RPL_i + \beta_2\, Pr\ oclivity_i + \beta_3\, PC_i + \beta_4\, RPLS_i + \beta_5\, RL_i + \beta_6\, SE_i + \varepsilon_i \tag{2}$$

$$Satisf_i = \beta_0 + \beta_1\, RPL_i + \beta_2\, Proclivity_i + \beta_3\, RPL_i\ X\ Proclivity_i + \beta_4\, PC_i + \beta_5\, RPLS_i + \beta_6\, RL_i + \beta_7\, SE_i + \varepsilon_i \tag{3}$$

where $Satisf_i$ indicates whether a review by user i has a positive or negative satisfaction. Model (2) examine the relationships between RP leniency dimensions on customer satisfaction where Model (3) examine the differences among these relationships moderating by three groups of *return proclivity*. We use the default parameters for both multi-nominal and binary logistic regression.

5. Results

5.1. The impacts of RP leniency on return proclivity

We start with the examination of RP leniency impacts on return proclivity. We pay attention to the odds ratio (φ) which measures the likelihood of an outcome category (e.g., return intention) occurring due to a specific exposure compared to the likelihood of it occurring without that exposure. When using a multinomial logistic model, odds ratios are easier to interpret in the context of categorical outcomes and offer a useful perspective for understanding the results (Sluis and De Giovanni, 2016). The odds ratios can be interpreted as the change in the odds of the outcome (e.g., $y = 1$) resulting from a one-unit increase in an independent variable and is expressed as $\varphi_i = e^{\beta i}$. If the odds ratio is less than 1, it indicates that the odds of being in that outcome category are lower when the leniency level is the level of interest (e.g., level 2) compared to when it is the reference level (e.g., level 1).

First, we apply a stepwise regression approach to systematically confirm the contribution of all control variables in Online Appendix F. Since all control variables improve the determination levels of the model, we use them as the base model. Table 7 displays the results for Model (1) in which each type of returners (prospective returners and returners) is compared to a reference non-returner. Model (1) is significantly better than the base model since the LR test is significant (p-value ≤0.01) according to Shang et al. (2017b). Furthermore, Model (1)'s AIC and Pseudo R2 are better than the base model – 17610.48 < 17755.38 and 0.2027 > 0.1941 respectively.

Time leniency dimension is not significant at any level for both customers with return intention and behaviour. *Monetary leniency* is negatively significant at level 1–3 and 2–3 for both return intention ($\beta = -0.159; \beta = -0.136$) and behaviour ($\beta = -0.317; \beta = -0.202$). This variable indicates that offering a full refund instead of charging fees or providing partial refunds would make customers less inclined to mention such experience in reviews (all $\varphi < 1$).

Effort leniency is significant at all levels, specifically level 1–2, 1–3 and 2–3 for prospective returners ($\beta = 0.213$; $\beta = -0.123$ and $\beta = -0.336$, respectively). This result only holds true for returners at level 1–2 and 2–3 ($\beta = 0.193$, and $\beta = -0.309$, respectively). In the return intention

Table 7
Multi nominal logistic regression for the impact of RP leniency dimensions on customer return proclivity.

	Return intention (reference = no intention)	Odd ratio, φ	Return behaviour (reference = no intention)	Odd ratio, φ
Intercept [a, b]	−3.752*** (0.171)		−4.499*** (0.220)	
T2 (reference = T1)	−0.022 (0.087)	0.978	−0.017 (0.112)	0.983
T3 (reference = T1)	−0.046 (0.103)	0.955	0.029 (0.131)	1.030
T3 (reference = T2)	−0.024 (0.073)	0.976	0.046 (0.091)	1.050
M2 (reference = M1)	−0.023 (0.080)	0.977	−0.115 (0.101)	0.892
M3 (reference = M1)	**−0.159** (0.066)**	0.853	**−0.317*** (0.083)**	0.729
M3 (reference = M2)	**−0.136* (0.076)**	0.873	**−0.202** (0.097)**	0.817
Ef2 (reference = Ef1)	**0.213*** (0.067)**	1.237	**0.193** (0.085)**	1.213
Ef3 (reference = Ef1)	**−0.123* (0.073)**	0.884	−0.115 (0.091)	0.891
Ef3 (reference = Ef2)	**−0.336*** (0.080)**	0.715	**−0.309*** (0.100)**	0.734
S2 (reference = S1)	**−0.343*** (0.067)**	0.709	**−0.271*** (0.084)**	0.763
S3 (reference = S1)	**−0.151* (0.078)**	0.860	−0.100 (0.098)	0.905
S3 (reference = S2)	**0.193*** (0.074)**	1.212	**0.171* (0.094)**	1.186
Ex2 (reference = Ex1)	0.063 (0.183)	1.065	**−0.512* (0.282)**	0.599
Ex3 (reference = Ex1)	**−0.131** (0.060)**	0.878	**−0.142* (0.075)**	0.867
Ex3 (reference = Ex2)	−0.193 (0.183)	0.824	0.370 (0.282)	1.447
RP Sentiment: Positive [a]	0.041 (0.059)	1.042	0.107 (0.074)	1.113
RP subjectivity [a]	−0.009 (0.140)	0.991	0.148 (0.175)	1.160
RP length [a]	0.001*** (0.000)	1.001	0.0003 (0.000)	1.000
Review length [a]	0.058*** (0.001)	1.060	0.060*** (0.001)	1.062
Product categories	Included		Included	
Seasonal effects	Included		Included	
N	25,037			
LLR	−8793.31			
LR Test [c]	80.75***			
AIC	17610.48			
Pseudo R2	0.2027			

*** p-value ≤0.01; ** p-value ≤0.05; * p-value ≤0.1.
[a] The coefficients are reported from the model where reference = level 1 for all leniency dimensions.
[b] Standard errors are in the parentheses.
[c] The LR Test row display the likelihood ratio test (χ^2 statistics) comparing between the log likelihoods of Model (1) and the base model with control variables only. Model 1 is significantly better than the base model as it is significant at 0.1.

group, customers are less inclined to mention their return willingness ($\varphi = 0.884$ and $\varphi = 0.715$) when e-tailers provide a hassle-free return process, compared to requiring original receipts, packaging, or forms (i.e., conditional returns). This trend is also seen in the return behaviour group at levels 2–3 ($\varphi = 0.734$) but not at levels 1–3. Past returners already returned when writing a review, indicating that the switch from conditional returns to a hassle-free does not affect their behaviour. Interestingly, if e-tailers switch from conditional returns to form-filling requirements, customers are more likely to express both return intentions and behaviours ($\varphi = 1.237$ and $\varphi = 1.213$).

Customers seem also be sensitive about *scope leniency* as they are statistical significance at all levels 1–2, 1–3 and 2–3 for prospective returners ($\beta = -0.343$; $\beta = -0.151$ and $\beta = 0.193$, respectively). This finding holds true for past returners only at level 1–2 and 2–3 ($\beta = -0.271$ and $\beta = 0.171$, respectively). If e-tailers slightly loosen their RP, such as allowing a specific product to be returned if it is unused instead of prohibiting its return entirely, customers are less likely to express return likelihood. Reversely, customers react more actively to the change from level 2 to 3 both return intentions and behaviours ($\varphi = 1.186$ and $\varphi = 1.212$, respectively). From levels 1–3 for returners, scope is not significant which may imply that their returned products are immune from the prohibiting list.

When *exchange leniency* is adjusted from level 1 to 2 and 3, the likelihood of return behaviour in reviews decrease significantly ($\beta = -0.512$ and $\beta = -0.142$, respectively). This only holds true for return intention at levels 1–3 ($\beta = -0.131$). By allowing cash refund instead of only allowing exchange/replace/repair, customers reduce their willingness to express their return proclivity ($\varphi = 0.878$ for prospective returners and $\varphi = 0.867$ for returners). Surprisingly, potential returners show no significant inclination ($p\text{-}value > 0.1$) towards being influenced by a minor shift in leniency, such as the option to exchange for store credit or gift cards, when compared to actual returners ($\varphi = 0.599$).

5.2. The impacts of RP leniency on satisfaction

Model (2) and Model (3) results display the impact of RP leniency and their interactions with return proclivity on customers satisfaction (Table 8). Similarly with return proclivity effects, we also apply a stepwise regression approach to systematically confirm the contribution of all control variables in Online Appendix F. Since all control variables improve the determination levels of the model, we use them as the base model. We also focus on the marginal effect (η) which quantifies the change in the probability of the dependent variable occurring for a one-unit change in an independent variable, while other variables remain constant. Since the effects of exchange leniency on the satisfaction are insignificant across all levels and type of returners, we focus more on the other dimensions. Model (2) & (3) are significantly better than the base model since the LR tests are significant (p-value ≤0.05) according to Shang et al. (2017b). Furthermore, Model (2) & (3)’s AIC and Pseudo R2 are better than the base model (i.e., 26557.20 & 26575.30 < 28305.86 and 0.176 & 0.177 > 0.1207 respectively).

Time leniency significantly impacts customers satisfaction at level 1–3 ($\beta = -0.116$). When e-tailers relax their RP from extremely short return window (e.g., 7–14 days) to very long return window (e.g., 60–90 days),

Table 8
Logistic regression for the impact of RP leniency dimensions on customer return satisfaction moderated by return proclivity.

	Model (2)	Model (3)	Marginal effect, η
Intercept [a, b]	1.542*** (0.104)	1.537*** (0.106)	
T2 (reference = T1)	−0.063 (0.048)	−0.063 (0.051)	−0.011
T3 (reference = T1)	**−0.116** (0.057)**	**−0.111* (0.060)**	−0.020
T3 (reference = T2)	−0.053 (0.040)	−0.048 (0.042)	−0.009
M2 (reference = M1)	0.032 (0.045)	0.047 (0.047)	0.008
M3 (reference = M1)	**0.071** (0.037)**	**0.080** (0.038)**	0.014
M3 (reference = M2)	0.042 (0.041)	0.034 (0.043)	0.006
Ef2 (reference = Ef1)	**0.068* (0.039)**	**0.077* (0.040)**	0.014
Ef3 (reference = Ef1)	−0.007 (0.040)	−0.013 (0.041)	−0.002
Ef3 (reference = Ef2)	−0.074 (0.046)	**−0.090* (0.048)**	−0.016
S2 (reference = S1)	0.052 (0.037)	0.046 (0.039)	0.008
S3 (reference = S1)	**0.124*** (0.044)**	**0.140*** (0.046)**	0.025
S3 (reference = S2)	**0.072* (0.041)**	**0.095** (0.043)**	0.017
Ex2 (reference = Ex1)	−0.073 (0.117)	−0.101 (0.122)	−0.018
Ex3 (reference = Ex1)	−0.026 (0.034)	−0.023 (0.035)	−0.004
Ex3 (reference = Ex2)	0.047 (0.116)	0.078 (0.121)	0.014
M2 (reference = M1) X Behaviour (reference = No Intention) [c]		**−0.386* (0.222)**	−0.069
Ef3 (reference = Ef1) X Behaviour (reference = No Intention) [c]		**0.359* (0.196)**	0.064
Ef3 (reference = Ef2) X Behaviour (reference = No Intention) [c]		**0.429** (0.204)**	0.077
S2 (reference = S1) X Intention (reference = No Intention) [c]		**0.328* (0.181)**	0.058
S3 (reference = S1) X Behaviour (reference = No Intention) [c]		**−0.415* (0.219)**	−0.074
Return proclivity: Intention [a]	−2.231*** (0.076)	−2.122*** (0.297)	−0.378
Return proclivity: Behaviour [a]	−1.096*** (0.075)	−0.968*** (0.294)	−0.173
RP Sentiment: Positive [a]	0.105*** (0.032)	0.103*** (0.032)	0.018
RP subjectivity [a]	0.269*** (0.080)	0.268*** (0.080)	0.048
RP length [a]	−0.000 (0.0001)	0.000 (0.0001)	4.98e-07
Review length [a]	−0.036*** (0.001)	−0.036*** (0.001)	−0.007
Product categories	Included	Included	
Seasonal effects	Included	Included	
N	25,037	25,037	
LLR	−13456.58	−13445.81	
LR Test [d]	23.93***	45.48**	
AIC	26557.20	26575.30	
R2	0.176	0.177	

*** p-value ≤0.01; ** p-value ≤0.05; * p-value ≤0.1.
[a] The coefficients are reported from the model where reference = level 1 for all leniency dimensions.
[b] Standard errors are in the parentheses.
[c] We examined all possible interaction terms but only reported the significant ones to keep the table more concise. The full interactions can be found in Online Appendix G.
[d] The LR Test row display the likelihood ratio test (χ^2 statistics) comparing the log likelihoods of Model (2) and (3) with the base model of control variables only. Model (2) and (3) are significantly better than the base model as it is significant at 0.1.

Table 9
Summary of main findings.

Leniency dimension	Level	Return proclivity		Satisfaction				Previous findings	
		Return intention	Return behaviour	Overall	No return intention	Return intention	Return behaviour	Dependent variables	Main Effects
Time	1–2							- Return proclivity (Janakiraman et al., 2016;	(−)
	2–3							Janakiraman and Ordóñez, 2012)	(+)
	1–3			(−)	(−)			- Service quality, purchased intention (Rao et al., 2018; Zhang et al., 2017)	
Monetary	1–2						(−)	- Return proclivity (Janakiraman et al., 2016;	No effect
	2–3	(−)	(−)					Wood, 2001; Zhang et al., 2017)	(+)
	1–3	(−)	(−)	(+)	(+)			- Product quality and satisfaction, customers spending (Bower and Maxham, 2012; Heim and Field, 2007; Wood, 2001)	
Effort	1–2	(+)	(+)	(+)	(+)			- Return proclivity/rate (Gäthke et al., 2022;	No
	2–3	(−)	(−)		(−)		(+)	Janakiraman et al., 2016)	effect/
	1–3	(−)					(+)	- Customer rating (Heim and Field, 2007)	(+)
									(+)
Scope	1–2	(−)	(−)			(+)		- Return proclivity (Janakiraman et al., 2016)	(+)
	2–3	(+)	(+)	(+)	(+)			- Perceive service quality (Abdulla et al., 2022)	(+)
	1–3	(−)		(+)	(+)		(−)		
Exchange	1–2		(−)					- Return proclivity (Janakiraman et al., 2016)	(−)
	2–3							- Customer rating (Heim and Field, 2007)	No effect
	1–3	(−)	(−)						

(+): The effect is positively significant at the level.
(−): The effect is negatively significant at the level.
Empty cell: The effect is not significant at the level.

customers tend to be more negative. This effect could be attributed by the non-returners as they also negatively influence customer satisfaction at level 1–3 in Model (3) ($\beta = -0.111$ and $\eta = -0.02$).

On the contrary, *monetary leniency* positively increases customers satisfaction at levels 1–3 ($\beta = 0.071$), especially in the no return intention group ($\beta = 0.080$ and $\eta = 0.014$). The result also indicates an unexpected effect in returners group. Customers decrease satisfaction when e-tailers change from partial refund to only shipping/restocking fees charged ($\beta = -0.386$ and $\eta = -0.069$).

Overall, *effort leniency* is only positively significant at level 1–2 ($\beta = 0.068$) which means asking to fill forms to contact them without the need of receipts/tag would increase customer satisfaction. However, when we moderated the effect with the return proclivity groups, it reveals some heterogeneous relationships. For the non-returner group, filling forms such as RMA is optimal as it has a positive impact from level 1 ($\beta = 0.077$ *and* $\eta = 0.014$) while having a negative impact if moving to level 3 ($\beta = -0.090$ *and* $\eta = -0.016$). For the returners group, a hassle-free policy marginally increases from level 1 and 2 customer satisfaction significantly ($\eta = 0.064$ and $\eta = 0.077$) while it has no effect on the prospective returners.

Scope leniency positively is significant at level 1–3 and 2–3 ($\beta = 0.124$ and $\beta = 0.072$, respectively). In other words, when e-tailers expand the list of returnable products or rule out the conditions of returned products, customers are happier with the e-tailer's services. These effects hold true for non-returners while it is not significant for prospective returners. We also found that the satisfaction of prospective returners increases significantly if the scope leniency moves from level 1 to level 2 ($\beta = 0.328$ *and* $\eta = 0.058$). Conversely, returners tend to express their experience negatively if e-tailers adjust the leniency from level 1 to level 3 ($\beta = -0.415$ *and* $\eta = -0.074$).

6. Discussions

Our research offers actionable insights for e-tailers to enhance return management through tailored RPs. Our findings are summarised in Table 9. We answer the research questions as follows.

6.1. RQ1. Can e-tailers issues RPs with restrictions without damaging their overall satisfaction?

- Time leniency dimension

A restricted RP can address the dilemma when it has higher or equal satisfaction compared with the permissive RP. We found mostly no significant impact of *time leniency* on return proclivity and satisfaction. This contrasts with the findings of Janakiraman et al. (2016); Janakiraman and Ordóñez (2012), who observed that a longer return window (e.g., from 2 to 7 days) discourages returns due to the endowment effect. Our results differed because we examined a broader range of return windows, from up to 14 days to over 30 days (e.g., 90 days). According to construal level theory, when the return deadline is distant, customers view the decision abstractly, whereas a shorter period prompts detailed attention to the return process and feasibility. With such extensive time windows, it is understandable that the propensity to return and satisfaction diminish in both prospective returners and past returners.

Additionally, time leniency significantly reduces satisfaction, especially among non-returners. This differs Rao et al. (2018) and Zhang et al. (2017), who found that a longer return window enhances perceived service quality and willingness to pay. A possible explanation is that an extended return period may heighten delayed gratification, causing customers to feel less urgency to use or assess the product. This delay can degrade customers' initial excitement and satisfaction, resulting in less positive reviews. Generally, we suggest that offering too generous return deadline can be unnecessary as it can backfire the marketplace e-tailers. A restricted return time window (e.g., 30 days) might be sufficient to keep customers happy.

- **Effort leniency dimension**

Our findings reveal that *effort leniency* required in RP has a varied impact on both return proclivity and customer satisfaction. Gäthke et al. (2022) observed that stringent effort requirements reduce return rates, while Janakiraman et al. (2016) found no such effect. In our study, effort leniency significantly increases returns proclivity and satisfaction at level 1–2. Interestingly, when e-tailers transition from requiring physical conditions (e.g., receipts, tags) to form-filling procedures (e.g., RMA), return proclivity and satisfaction positively increase in a review. Contrary to traditional views that RMAs are the least lenient RPs (Li et al., 2011), customers may find retaining receipts or original packaging more cumbersome than completing forms. Different with Heim and Field (2007), a hassle-free policy significantly decreases returns proclivity while not significant for satisfaction. This demonstrates that a hassle-free does not add much positive value to customer perception compared with form-filling procedures. This only leads to reduce controversy from the customers' perspective, thereby diminishing their interest in sharing their return experience. This finding implies the potential of employing form-filling/contact procedures (e.g., RMA) to address the dilemma.

- **Exchange leniency dimension**

Regarding *exchange leniency*, our finding suggests that customers are less likely to express their return proclivity if offered cash refund or gift card compared with exchange/repair (level 1–2 and 1–3). We found no effect of this dimension on customer overall satisfaction. These findings are in line with Janakiraman et al. (2016) and Heim and Field (2007). Janakiraman et al. (2016) explained that the negative direction of exchange leniency on return proclivity is because consumers with minor product problems are more likely to utilise exchanges when such options are appealing. Hence, we suggest that promoting exchange/repair while highlighting unique selling points would arouse customers' attachment with the products/brand, leading to the preference towards replacement/repair.

- **Monetary leniency dimension**

Previous literature found no effect between *monetary leniency* and return proclivity (Janakiraman et al., 2016; Wood, 2001; Zhang et al., 2017). Surprisingly, we found that transitioning to full refund (level 2–3, 1–3) significantly reduces return proclivity. Our finding also demonstrates that full refund would significantly enhance customer satisfaction at level 1–3 which is in line with Bower and Maxham (2012); Heim and Field (2007) and Wood (2001). However, the widespread availability of free returns, often bundled with free shipping, has reshaped customer expectations, making full refunds (e.g., no shipping fee from Amazon) an industry norm rather than a distinguishing benefit. According to customer delight theory (Arnold et al., 2005), exceeding expectations is key to delighting customers. However, when free returns become expected, they may no longer generate strong positive reactions, leading to reduced discussion in reviews. This shift in expectations may explain why we observe a decrease in return proclivity despite greater monetary leniency. Customers who receive full refunds may not feel compelled to mention it unless faced with unexpected restrictions, making the reduction in return proclivity a potential false signal. As customers are generally more satisfied at level 1–3, we find no acceptable restricted level to address the dilemma in this dimension.

- **Scope leniency dimension**

Scope leniency also has a heterogeneous effect on return proclivity and satisfaction. When e-tailers ease their policies at levels 1–2, customers are less likely to initiate returns proclivity. Our finding contrasts with Janakiraman et al. (2016) who found that restrictiveness compared with no restrictions in the product scope would dissuade return tendency. In our context, a relax from total to conditional prohibition might create more troubles for customers. For example, furniture must be disassembled and re-packaged according to Walmart. At level 2–3, we observed significant increases in satisfaction and likelihood of return expression which in line with Abdulla et al. (2022) and Janakiraman et al. (2016). This is logical, as keeping products unused or unopened is difficult; thus, customers would praise e-tailers who eliminate this constraint. Therefore, we do not find an acceptable policy for this dimension.

6.2. RQ2. Do RP leniency dimensions have different effects among return intention and behaviour?

Another noteworthy observation is that satisfaction of both prospective returners and returners has not almost been affected by *monetary leniency*. Satisfaction is only negatively significant at level 1–2 for customers who have returned. This can be explained as customers might irrationally weigh the initial cost of a purchase heavily. Any refund, regardless the amount, can be perceived as a gain, leading to similar satisfaction levels. This phenomenon is known as sunk cost fallacy (Arkes and Blumer, 1985). Prospective/Past returners might accept monetary restrictions at level 1–2, only if e-tailers can highlight the heavy weight of initial purchasing costs.

In terms of *effort leniency*, different with our previous finding in proposing medium level – form-filling procedures (e.g., RMA), we found that hassle-free is an ideal policy for return behaviour group. This is since customers with prior returns expressed positive sentiments towards the hassle-free policy, aligning with Heim and Field (2007). However, this policy does not affect the satisfaction of prospective/non returners, as they have not yet engaged with the return process. For e-tailers aiming to boost rating where prospective/non returners predominate, incorporating a minor hassle such as an RMA could be beneficial. Conversely, a hassle-free policy would be more advantageous for retaining the satisfaction and loyalty of returners.

Although we found no optimal level for *scope leniency* to address the dilemma, it is intriguing that prospective returners increase their satisfaction significantly when offered conditional returns. These customers may be aware that these products are non-returnable such as hazardous or bulky furniture. If they are allowed to be returned, regardless of conditions, this group sees this policy positively. Furthermore, past/ongoing returners experience downward satisfaction at level 1–3. Customers might feel awkward about returning items perceived as non-standard and the logistics of returning a bulky or hazardous item can amplify discomfort. This offers valuable insights for e-tailers. We analyse Walmart's restricted items list in which prohibitions are due to safety concerns, aligning with governmental guidelines. E-tailers should follow Walmart by either presenting compelling rationales for item prohibitions or offering explicit return guidelines under specified conditions.

6.3. Robustness check

In this section, we strengthen the robustness of our analysis by employing two models: Structural Equation Modelling (SEM) and Random Forest (RF), to examine the relationships between the five RP leniency dimensions and both return proclivity and satisfaction.

First, we apply SEM, which offers a comprehensive view of how observed and latent variables interact, while assessing the overall model fit. This approach ensures that the relationships identified in logistic regression hold within a more sophisticated framework that accounts for errors. Our SEM results confirm that the significance levels and directions remain consistent with those observed in the main models. Furthermore, we found no significant reverse causality between RPs leniency dimensions and the dependent variables—return proclivity and satisfaction. This suggests that endogeneity is not a major concern in our models. Detailed SEM results are provided in Online Appendix H1.

Next, to account for non-linear relationships between the RP leniency dimensions and the two dependent variables, we use RF to validate the patterns identified in the logistic regression, following the methodology outlined by Cui et al. (2020). The goodness-of-fit metrics for the RF models consistently exceed 0.7 across three common indicators: AUC, Precision, and Recall. Since the relationships are non-linear and require further interpretability, we also use Partial Dependence Plots (PDPs) (Friedman, 2001). PDPs are well-suited for assessing how return proclivity and satisfaction depend on the RP leniency dimensions by varying the value of the examined variables while holding others constant—similar to the use of control variables in logistic regression. The PDP results demonstrate consistent non-linear relationships, aligning with the findings from logistic regression. Detailed RF results can be found in Online Appendix H2.

7. Implications and contributions

7.1. Theoretical implications and contributions

This study advances two major domains: (1) customer return behaviour research, and (2) empirical RP classification frameworks, by offering new insights into how legitimate consumers respond to varying levels of leniency across five RP dimensions—time, monetary, effort, scope, and exchange—through their return intention, behaviour, and satisfaction.

First, our study is among the first to formally integrate Expectancy Disconfirmation Theory, Signalling Theory, and Justice Theory into a unified framework to explain consumer reactions to RP leniency, potentially address the *"Return policy leniency dilemma"*. In contrast to prior research that often relies on binary classifications (e.g., lenient vs strict), we found that moderately restrictive policies – specifically, 30-day return windows (time), requiring RMA (effort), and offering exchange/repair instead of cash refunds (exchange) – are preferred by legitimate customers, challenging the dominant assumption that more leniency always results in higher satisfaction. This finding fills a critical theoretical gap by highlighting that customer satisfaction may not follow a simple linear relationship with leniency.

Second, we extend the RP classification literature by introducing a three-tiered leniency categorisation (low, medium, high) across all five dimensions. Previous studies either focus on a single or partial set of dimensions or group low and medium leniency together. By disaggregating these levels using real-world policy data, our study offers a more nuanced and empirically grounded framework to analyse consumer responses. This enriched classification allows for a deeper understanding of how subtle variations in RP design affect consumer behaviour and satisfaction—offering practical guidance for e-retailers seeking to optimise return strategies.

Third, we shift the focus from fraudulent return behaviour – the common trend in literature – to the perspectives of legitimate customers, who constitute the majority of e-commerce shoppers. Prior work has often overlooked how restrictive RPs are perceived by honest consumers. Our findings show that satisfaction levels vary significantly between non-returners, prospective returners, and past returners, particularly in response to monetary, effort, and scope-related policies. This insight addresses a critical deficiency in the literature and offers actionable recommendations for small and medium-sized e-tailers aiming to reduce returns without alienating their customer base.

7.2. Practical implications and contributions

Our study also makes some practical implications and contributions to e-tailers. First, our comprehensive framework serves as a strategic guide for e-tailers to navigate and position themselves within the three levels (low, medium, and high) of the five RP dimensions. This enables businesses to project the perceptions of legitimate customers – including non-returners, prospective returners, and past returners – and to tailor their RPs accordingly. By leveraging this framework, e-tailers can strike a balance between return satisfaction and minimising return proclivity, achieving greater alignment with customer expectations while addressing the RP leniency dilemma.

More specifically, SMEs across different industries can strategically tailor their RPs to balance customer satisfaction with operational efficiency. For example, fashion retailers may benefit from moderate exchange and monetary leniency, such as offering replacements instead of full refunds and charging small restocking fees to manage frequent size/fit-related returns. Electronics retailers could adopt stricter time leniency, providing a reasonable return window (e.g., 30 days) that allows customers to assess product quality (e.g., durability) without extending it so long that resale value declines or fraud risks increase. Meanwhile, consumable goods retailers (e.g., food and supplements) may rely on effort-based restrictions, such as requiring return authorisation forms, to deter unnecessary returns while ensuring fairness for legitimate claims. By carefully adjusting RPs based on industry-specific challenges, SMEs can enhance customer trust while minimising return-related costs and risks.

Second, our machine learning-based text mining approach offers actionable insights by analysing a large textual dataset of real-world RPs and customer reviews. This method provides a holistic perspective on how e-tailers craft their RPs, including a wide range of terms and conditions and their nuanced relationships with return proclivity and satisfaction. These insights empower new and small businesses to strategically design RPs that maximise resources while effectively managing product returns. Additionally, this approach demonstrates the practical utility of customer language analysis, revealing how customers perceive and react to RPs based on the language used in policy descriptions and their personal experiences.

8. Conclusion

With the growth of e-commerce, RP leniency management have become more crucial to customer satisfaction since they are more inclined to seek product assurance to mitigate the prepurchase product quality uncertainties. However, this also means that their online return intention and action can be escalated through this ease of return. This refers to the RP leniency dilemma. To understand and provide guidance on this, our study shed new light by examining the effects of the five RP leniency dimensions on customer return proclivity and satisfaction.

We found that acceptable/better customer satisfaction while lower customer return proclivity cannot be achieved by highly permissive RP. Our central insights exhibit more complicated relationships, furnishing tailored guidance for e-tailers. With that in mind, the practical contribution of this study is providing e-tailers with bespoke RP designs incorporating all five leniency dimensions that can help them have a better chance of addressing the RP leniency dilemma. Depending on the current situation and where each e-tailer want to be, our findings empower them to self-assess and form their own strategy systematically. We contribute to enhance return management through tailored RP for e-tailers.

This study is also subject to certain limitations. First, this study focuses on Walmart dataset, subsequent research endeavours should investigate methodologies for extracting e-tailer RPs across prominent platforms such as Amazon, eBay, or Alibaba, aiming to provide a more comprehensive perspective. Second, we have used product categories as a control variable to ensure that generality of our results. However, future research may look at how the customers' perspective differs among distinct product categories. Third, this study does not focus on temporal changes in RP leniency or seasonal fluctuations in return behaviour. While our cross-sectional approach provides a generalised view across multiple retailers, a longitudinal study focusing on how a single e-tailer's policy changes over time might offer additional insights into evolving consumer response. Fourth, we assume the customers are fully aware of RP. As such, our findings may not generalise to customers who are less aware of, or indifferent to, RP details. Although this subset offers rich insights into policy-aware customer behaviour, future research could explore the role of policy awareness and its timing (pre- vs. post-purchase) using alternative data sources or experimental designs. Lastly, a key limitation of this study is the reliance on seller ratings as a proxy for seller reputation, which may not capture the full spectrum of reputation, particularly given that platforms like Walmart typically restrict participation to higher reputation e-tailers. While previous research suggests both seller reputation and RP leniency influence each other (Robertson et al., 2020; Walsh et al., 2016), our dataset lacks sufficient variation in seller reputation to fully address this confounding factor. Future research could employ experimental methods or additional reputation measures (e.g., Amazon sales rank) to better isolate the effects of seller reputation on return behaviour and satisfaction. This highlights the broader limitation of data-driven approaches compared to experimental designs, which allow for more direct manipulation and control over confounding variables.

CRediT authorship contribution statement

Quang Huy Duong: Writing – review & editing, Writing – original draft, Visualization, Validation, Resources, Methodology, Formal analysis, Data curation, Conceptualization. **Li Zhou:** Writing – review & editing, Validation, Supervision, Methodology, Formal analysis, Data curation, Conceptualization. **Meng Meng:** Writing – review & editing, Writing – original draft, Validation, Supervision, Methodology, Formal analysis, Conceptualization. **Le Thuy An Dang:** Writing – review & editing, Writing – original draft, Formal analysis. **Tiep Duy Nguyen:** Writing – original draft, Methodology, Formal analysis, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Appendix I. Variable description, VIF and correlations

	Variables	Mean	Std	VIF	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
1	Return Proclivity: Intention	0.07	0.26	1.34	1.0																
2	Return Proclivity: Behaviour	0.04	0.2	1.19	−0.06***	1.0															
3	Satisfaction: Positive	0.66	0.47	3.25	−0.32***	−0.17***	1.0														
4	T1	0.13	0.33	1.24	−0.01*	−0.01*	0.03***	1.0													
5	T3	0.18	0.38	1.27	−0.0	−0.0	−0.01	−0.18***	1.0												
6	M1	0.36	0.48	2.07	0.03***	0.03***	−0.03***	−0.05***	0.0	1.0											
7	M2	0.22	0.41	1.63	0.02*	0.02*	−0.03***	−0.06***	−0.06***	−0.4***	1.0										
8	Ef2	0.31	0.46	1.98	0.02**	0.01	0.02**	0.09***	−0.05***	0.2***	−0.18***	1.0									
9	Ef3	0.31	0.46	2.11	0.0	0.01	−0.04***	−0.11***	0.01	−0.26***	0.27***	−0.45***	1.0								
10	S1	0.32	0.47	1.93	0.03***	0.02***	−0.04***	−0.0	0.01	−0.15***	0.23***	−0.16***	0.24***	1.0							
11	S3	0.2	0.4	1.49	−0.01	−0.01	0.03***	0.05***	0.08***	−0.08***	−0.07***	−0.0	0.02**	−0.35***	1.0						
12	Ex1	0.4	0.49	1.98	0.02**	0.01*	0.01	0.04***	−0.01	0.13***	−0.17***	0.22***	−0.21***	−0.08***	−0.02***	1.0					
13	Ex2	0.02	0.13	1.05	0.02*	−0.0	−0.01*	−0.02***	0.03***	0.07***	−0.03***	0.01*	−0.04***	−0.02**	0.04***	−0.11***	1.0				
14	RP Subjectivity	0.48	0.19	6.00	0.03***	0.03***	−0.01	−0.02***	−0.05***	−0.02**	−0.03***	−0.05***	−0.05***	−0.08***	0.05***	−0.0	0.01	1.0			
15	RP Sentiment: Positive	0.65	0.48	2.84	0.02*	0.02**	0.0	−0.0	0.02***	−0.08***	0.0	−0.0	0.05***	0.13***	0.05***	−0.09***	0.01	0.04***	1.0		
16	RP Length	188.97	167.45	3.49	0.02***	0.01	0.01	0.12***	−0.06***	0.35***	−0.28***	0.39***	−0.46***	−0.24***	−0.02***	0.36***	0.05***	0.0	−0.11***	1.0	
17	Review Length	23.45	19.27	3.26	0.35***	0.28***	−0.38***	−0.06***	−0.0	0.03***	0.06***	−0.04***	0.09***	0.05***	−0.02***	−0.03***	0.02**	0.08***	0.03***	−0.04***	1.0
18	Category: Electronics	0.14	0.35	1.53	0.05***	0.04***	−0.05***	−0.03***	0.02***	0.04***	0.07***	0.03***	0.04***	0.06***	0.01	0.05***	−0.02**	0.1***	0.03***	0.0	0.1***
19	Category: Furniture and Furnishing	0.19	0.39	1.68	−0.03***	−0.01	0.04***	0.1***	−0.05***	−0.07***	0.02*	0.1***	0.02**	0.1***	−0.04***	0.01*	−0.04***	0.06***	−0.01*	0.05***	−0.03***
20	Category: Grocery and Food	0.04	0.21	1.14	−0.02**	−0.02*	0.01	−0.04***	−0.04***	−0.06***	0.04***	−0.03***	0.06***	0.0	−0.02**	−0.07***	−0.01	−0.08***	−0.03***	−0.07***	−0.01
21	Category: Health, Personal Care and Beauty	0.1	0.3	1.31	−0.01	−0.01	−0.02**	−0.05***	−0.01*	0.01	−0.01	−0.02**	0.06***	−0.06***	−0.0	−0.05***	0.03***	−0.02**	−0.04***	−0.07***	−0.0
22	Category: Recreation, Sport and Culture	0.18	0.39	1.56	−0.03***	−0.03***	0.02***	−0.02*	0.03***	0.06***	0.0	−0.02**	−0.03***	−0.07***	−0.01	−0.0	0.01	−0.08***	−0.04***	0.06***	−0.04***
23	Category: Vehicles	0.04	0.19	1.13	0.0	−0.01	0.0	0.03***	0.01	0.04***	−0.0	0.01	−0.01*	0.0	0.0	0.04***	−0.02***	−0.01*	−0.06***	−0.02**	−0.02*
24	Month: January	0.12	0.32	1.7	0.01	−0.0	0.01*	−0.01	−0.0	0.01	0.0	0.03***	−0.02*	−0.0	−0.01	0.02*	0.0	0.01	0.02***	0.02*	0.01
25	Month: February	0.11	0.31	1.65	0.0	0.01*	0.02**	−0.02***	−0.02*	0.01	0.0	−0.0	−0.0	−0.01*	0.01*	−0.01	−0.0	−0.01	0.01	−0.01	−0.02**
26	Month: March	0.1	0.3	1.58	0.0	0.0	0.0	−0.03***	0.01	0.01	0.01	−0.01	0.02***	0.02*	0.0	−0.02**	0.0	−0.02*	0.0	−0.04***	−0.03***
27	Month: April	0.09	0.29	1.54	0.01	−0.0	−0.0	−0.01	0.01	−0.01	0.01	−0.04***	0.04***	0.0	0.0	−0.03***	−0.02*	−0.01	−0.02***	−0.05***	0.02**
28	Month: May	0.08	0.27	1.46	0.0	−0.01*	−0.01	−0.01*	0.03***	−0.0	0.01	−0.03***	0.03***	0.01*	−0.01	−0.03***	−0.0	−0.0	−0.01	−0.05***	0.02**
29	Month: June	0.05	0.23	1.33	0.01	0.01	−0.02**	−0.02**	0.01	−0.01	0.01	−0.04***	0.03***	0.0	0.01	−0.01	0.0	0.02*	−0.02***	−0.03***	0.04***
30	Month: July	0.06	0.24	1.36	−0.01	0.01	−0.02**	−0.01	0.0	−0.0	0.01	−0.04***	0.03***	0.0	0.01	−0.03***	0.0	−0.01	−0.01	−0.04***	0.02***
31	Month: August	0.08	0.27	1.44	−0.01*	−0.02***	−0.0	0.02**	0.01	−0.0	−0.01	−0.01	−0.0	0.01	0.01	0.01	0.01	−0.02*	−0.01	−0.0	−0.01
32	Month: September	0.06	0.24	1.36	−0.02**	0.0	0.01*	0.02***	−0.01	−0.02**	0.0	0.02**	−0.02***	0.0	0.01*	0.03***	0.0	0.01	0.01	0.01*	−0.02***
33	Month: October	0.06	0.24	1.38	−0.01	0.01	0.01	0.04***	−0.03***	0.02**	−0.01*	0.02**	−0.02***	0.0	−0.02**	0.02***	−0.01	0.02***	0.01*	0.05***	0.0
34	Month: November	0.07	0.25	1.41	−0.02**	−0.02*	0.01*	0.02***	−0.01	0.02*	−0.01	0.02**	−0.02***	−0.01	−0.01*	0.01	−0.0	0.01	0.0	0.04***	−0.01

	Variables	**18**	**19**	**20**	**21**	**22**	**23**	**24**	**25**	**26**	**27**	**28**	**29**	**30**	**31**	**32**	**33**	**34**
18	Category: Electronics	1.0																
19	Category: Furniture and Furnishing	−0.2***	1.0															
20	Category: Grocery and Food	−0.09***	−0.1***	1.0														
21	Category: Health, Personal Care and Beauty	−0.13***	−0.16***	−0.07***	1.0													
22	Category: Recreation, Sport and Culture	−0.19***	−0.23***	−0.1***	−0.15***	1.0												
23	Category: Vehicles	−0.08***	−0.09***	−0.04***	−0.06***	−0.09***	1.0											
24	Month: January	0.02**	−0.02***	−0.02***	−0.01	0.0	−0.02*	1.0										
25	Month: February	0.01	−0.02**	0.0	0.0	0.01	−0.01	−0.13***	1.0									

(*continued on next page*)

(continued)

Variables	Mean	Std	VIF	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
26 Month: March				0.01	−0.01	0.02**	0.02***	−0.01	0.01	−0.12***	−0.12***	1.0								
27 Month: April				0.0	−0.0	0.02***	0.02**	−0.01	0.0	−0.11***	−0.11***	−0.1***	1.0							
28 Month: May				−0.0	0.01	0.0	0.02***	−0.01	0.02**	−0.1***	−0.1***	−0.1***	−0.09***	1.0						
29 Month: June				−0.0	−0.0	0.01	0.02**	−0.01	0.01	−0.09***	−0.08***	−0.08***	−0.07***	−0.07***	1.0					
30 Month: July				−0.01	−0.02**	0.0	0.02**	0.0	0.02*	−0.09***	−0.09***	−0.08***	−0.08***	−0.07***	−0.06***	1.0				
31 Month: August				0.01	0.0	0.01	−0.0	−0.0	0.01	−0.1***	−0.1***	−0.09***	−0.09***	−0.08***	−0.07***	−0.07***	1.0			
32 Month: September				0.0	0.02*	−0.0	−0.01	−0.01	0.01	−0.09***	−0.09***	−0.08***	−0.08***	−0.07***	−0.06***	−0.06***	−0.07***	1.0		
33 Month: October				−0.02***	0.01	−0.01	−0.02***	0.01	−0.0	−0.09***	−0.09***	−0.08***	−0.08***	−0.07***	−0.06***	−0.06***	−0.07***	−0.06***	1.0	
34 Month: November				−0.01	0.01	−0.01*	−0.01	0.01	−0.0	−0.1***	−0.1***	−0.09***	−0.08***	−0.08***	−0.06***	−0.07***	−0.08***	−0.07***	−0.07***	1.0

Number of observations: 25,037 reviews.
** Correlation significant at level 0.01 (2-tailed), * Correlation significant at level 0.05 (2-tailed).

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.jretconser.2025.104315.

Data availability

Data will be made available on request.

References

Abdulla, H., Abbey, J.D., Ketzenberg, M., 2022. How consumers value retailer's return policy leniency levers: an empirical investigation. Prod. Oper. Manag. 31, 1719–1733. https://doi.org/10.1111/POMS.13640.

Abdulla, H., Ketzenberg, M., Abbey, J.D., 2019. Taking stock of consumer returns: a review and classification of the literature. J. Oper. Manag. 65, 560–605. https://doi.org/10.1002/joom.1047.

An, Y., Kim, D., Lee, J., Oh, H., Lee, J.-S., Jeong, D., 2023. Topic modeling-based framework for extracting marketing information from E-commerce reviews. IEEE Access 11, 135049–135060. https://doi.org/10.1109/ACCESS.2023.3337808.

Anderson, R.E., 1973. Consumer dissatisfaction: the effect of disconfirmed expectancy on perceived product performance. J. Mark. Res. 10, 38. https://doi.org/10.2307/3149407.

Arkes, H.R., Blumer, C., 1985. The psychology of sunk cost. Organ. Behav. Hum. Decis. Process. 35, 124–140. https://doi.org/10.1016/0749-5978(85)90049-4.

Arnold, M.J., Reynolds, K.E., Ponder, N., Lueg, J.E., 2005. Customer delight in a retail context: investigating delightful and terrible shopping experiences. J. Bus. Res. 58. https://doi.org/10.1016/j.jbusres.2004.01.006.

Arts, J.W.C., Frambach, R.T., Bijmolt, T.H.A., 2011. Generalizations on consumer innovation adoption: a meta-analysis on drivers of intention and behavior. Int. J. Res. Market. 28, 134–144. https://doi.org/10.1016/j.ijresmar.2010.11.002.

Barbieri, F., Camacho-Collados, J., Espinosa Anke, L., Neves, L., 2020. TweetEval: unified benchmark and comparative evaluation for tweet classification. In: Findings of the Association for Computational Linguistics: EMNLP 2020. Association for Computational Linguistics, Stroudsburg, PA, USA, pp. 1644–1650. https://doi.org/10.18653/v1/2020.findings-emnlp.148.

Bicalho, P., Pita, M., Pedrosa, G., Lacerda, A., Pappa, G.L., 2017. A general framework to expand short text for topic modeling. Inf. Sci. 393, 66–81. https://doi.org/10.1016/j.ins.2017.02.007.

Bonifield, C., Cole, C., Schultz, R.L., 2010. Product returns on the Internet: a case of mixed signals? J. Bus. Res. 63, 1058–1065. https://doi.org/10.1016/J.JBUSRES.2008.12.009.

Bower, A.B., Maxham, J.G., 2012. Return shipping policies of online retailers: normative assumptions and the long-term consequences of fee and free returns. J. Market. 76, 110–124. https://doi.org/10.1509/jm.10.0419.

Chandrashekaran, M., Rotte, K., Tax, S.S., Grewal, R., 2007. Satisfaction strength and customer loyalty. J. Mark. Res. 44, 153–163. https://doi.org/10.1509/jmkr.44.1.153.

Chatterjee, A., Chauradia, A.J., Pedada, K., 2024. Rural women microentrepreneurs, consumer acquisition, and value delivery: evidence from a quasi-experiment in rural India. J. Acad. Market. Sci. https://doi.org/10.1007/S11747-024-01053-6.

Chen, J., Yu, B., Chen, B., Liu, Z., 2023. Lenient vs. stringent returns policies in the presence of fraudulent returns: the role of customers' fairness perceptions. Omega 117, 102843. https://doi.org/10.1016/j.omega.2023.102843.

Chen, X., Tan, L., Wang, F., 2024. Managing consumer bracketing purchases: optimal return policy analysis. Int. J. Prod. Econ. 275, 109321. https://doi.org/10.1016/j.ijpe.2024.109321.

Chen, Y., Wang, T., Zhu, S., Lian, P., 2020. Will you miss me if I am leaving? Unexpected market withdrawal of Norwegian Joy and customer satisfaction. Tour. Manag. 76, 103951. https://doi.org/10.1016/j.tourman.2019.103951.

Cohen, J., 1960. A coefficient of agreement for nominal scales. Educ. Psychol. Meas. 20, 37–46. https://doi.org/10.1177/001316446002000104.

Cui, H., Rajagopalan, S., Ward, A.R., 2020. Predicting product return volume using machine learning methods. Eur. J. Oper. Res. 281, 612–627. https://doi.org/10.1016/j.ejor.2019.05.046.

Dailey, L.C., Ülkü, M.A., 2018. Retailers beware: on denied product returns and consumer behavior. J. Bus. Res. 86, 202–209. https://doi.org/10.1016/j.jbusres.2018.01.064.

Davis, S., Gerstner, E., Hagerty, M., 1995. Money back guarantees in retailing: matching products to consumer tastes. J. Retailing 71, 7–22. https://doi.org/10.1016/0022-4359(95)90010-1.

de Mesquita, J.M.C., Shin, H., Urdan, A.T., Pimenta, M.T.C., 2023. Measuring the intention-behavior gap in service failure and recovery: the moderating roles of failure severity and service recovery satisfaction. Eur. J. Market. 57. https://doi.org/10.1108/EJM-03-2022-0235.

Desiraju, K., Mishra, A.N., Sengupta, P., 2024. Customer perceptions on open banking apps: insights using Structural topic modeling. J. Retailing Consum. Serv. 81, 104029. https://doi.org/10.1016/J.JRETCONSER.2024.104029.

Devlin, J., Chang, M.W., Lee, K., Toutanova, K., 2019. BERT: pre-training of deep bidirectional transformers for language understanding. In: NAACL HLT 2019 - 2019

Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies - Proceedings of the Conference.
Dickler, J., 2022. Free returns may soon be a thing of the past as retailers roll out stricter policies [WWW Document]. CNBC. URL. https://www.cnbc.com/2022/11/04/free-returns-may-be-over-as-retailers-introduce-stricter-policies.html.
Duong, Q.H., Zhou, L., Meng, M., Nguyen, T. Van, Ieromonachou, P., Nguyen, D.T., 2022. Understanding product returns: a systematic literature review using machine learning and bibliometric analysis. Int. J. Prod. Econ. 243. https://doi.org/10.1016/j.ijpe.2021.108340.
Duong, Q.H., Zhou, L., Van Nguyen, T., Meng, M., 2024. Understanding and predicting online product return behavior: an interpretable machine learning approach. Int. J. Prod. Econ., 109499 https://doi.org/10.1016/J.IJPE.2024.109499.
Finch, B.J., 2007. Customer expectations in online auction environments: an exploratory study of customer feedback and risk. J. Oper. Manag. 25. https://doi.org/10.1016/j.jom.2006.10.007.
Fornell, C., 1992. A national customer satisfaction barometer: the Swedish experience. J. Market. 56. https://doi.org/10.1177/002224299205600103.
Friedman, J.H., 2001. Greedy function approximation: a gradient boosting machine. Ann. Stat. 29. https://doi.org/10.1214/aos/1013203451.
Gäthke, J., Gelbrich, K., Chen, S., 2022. A cross-national service strategy to manage product returns: E-tailers' return policies and the legitimating role of the institutional environment. J. Serv. Res. 25. https://doi.org/10.1177/1094670521989440.
Greenberg, J., 1990. Organizational justice: yesterday, today, and tomorrow. J. Manag. 16, 399–432. https://doi.org/10.1177/014920639001600208.
Grootendorst, M., 2022. BERTopic: neural topic modeling with a class-based TF-IDF procedure. https://doi.org/10.48550/arxiv.2203.05794.
Heim, G.R., Field, J.M., 2007. Process drivers of e-service quality: analysis of data from an online rating site. J. Oper. Manag. 25, 962–984. https://doi.org/10.1016/J.JOM.2006.10.002.
Hjort, K., Lantz, B., 2016. The impact of returns policies on profitability: a fashion e-commerce case. J. Bus. Res. 69, 4980–4985. https://doi.org/10.1016/j.jbusres.2016.04.064.
Hsiao, L., Chen, Y.J., 2014. Return policy: hassle-free or your money-back guarantee? Nav. Res. Logist. 61, 403–417. https://doi.org/10.1002/nav.21591.
Janakiraman, N., Ordóñez, L., 2012. Effect of effort and deadlines on consumer product returns. J. Consum. Psychol. 22, 260–271. https://doi.org/10.1016/j.jcps.2011.05.002.
Janakiraman, N., Syrdal, H.A., Freling, R., 2016. The effect of return policy leniency on consumer purchase and return decisions: a meta-analytic review. J. Retailing 92, 226–235. https://doi.org/10.1016/J.JRETAI.2015.11.002.
Ketzenberg, M.E., Abbey, J.D., Heim, G.R., Kumar, S., 2020. Assessing customer return behaviors through data analytics. J. Oper. Manag. 66, 622–645. https://doi.org/10.1002/JOOM.1086.
Ko, E.E., Bowman, D., 2023. Suspicious online product reviews: an empirical analysis of brand and product characteristics using Amazon data. Int. J. Res. Market. 40, 898–911. https://doi.org/10.1016/J.IJRESMAR.2023.06.006.
Li, K.J., Fong, D.K.H., Xu, S.H., 2011. Managing trade-in programs based on product characteristics and customer heterogeneity in business-to-business markets. Manuf. Serv. Oper. Manag. 13, 108–123. https://doi.org/10.1287/msom.1100.0307.
Loeb, W., 2022. Free Internet returns are costly - will consumers have to pay? [WWW Document]. Forbes. URL. https://www.forbes.com/sites/walterloeb/2022/05/23/free-internet-returns-are-costlywill-consumers-have-to-pay/.
Menke, W.W., 1969. Determination of warranty reserves. Manag. Sci. 15, 542–549. https://doi.org/10.1287/mnsc.15.10.b542.
Minnema, A., Bijmolt, T.H.A., Gensler, S., Wiesel, T., 2016. To keep or not to keep: effects of online customer reviews on product returns. J. Retailing 92, 253–267. https://doi.org/10.1016/j.jretai.2016.03.001.
Mollenkopf, D.A., Frankel, R., Russo, I., 2011. Creating value through returns management: exploring the marketing-operations interface. J. Oper. Manag. 29, 391–403. https://doi.org/10.1016/j.jom.2010.11.004.
Mudambi, S.M., Schuff, D., 2010. What makes a helpful online review? A study of customer reviews on amazon.com. MIQ. https://doi.org/10.2307/20721420.
Mudambi, S.M., Schuff, D., Zhang, Zhewei, 2014. Why aren't the stars aligned? An analysis of online review content and star ratings. In: 2014 47th Hawaii International Conference on System Sciences. IEEE, pp. 3139–3147. https://doi.org/10.1109/HICSS.2014.389.
Oghazi, P., Karlsson, S., Hellström, D., Hjort, K., 2018. Online purchase return policy leniency and purchase decision: mediating role of consumer trust. J. Retailing Consum. Serv. 41, 190–200. https://doi.org/10.1016/J.JRETCONSER.2017.12.007.
Oliver, R.L., 1980. A cognitive model of the antecedents and consequences of satisfaction decisions. J. Mark. Res. 17, 460–469. https://doi.org/10.1177/002224378001700405.
Patankar, J.G., Mitra, A., 1995. Effects of warranty execution on warranty reserve costs. Manag. Sci. 41. https://doi.org/10.1287/mnsc.41.3.395.
Pei, Z., Paswan, A., Yan, R., 2014. E-tailer's return policy, consumer's perception of return policy fairness and purchase intention. J. Retailing Consum. Serv. 21. https://doi.org/10.1016/j.jretconser.2014.01.004.
Petro, G., 2022. Retailers risk alienating shoppers with stricter return policies. Forbes [WWW Document]. https://www.forbes.com/sites/gregpetro/2022/12/10/retailers-risk-alienating-shoppers-with-stricter-return-policies/.
Qiu, J., Liu, C., Li, Y., Lin, Z., 2018. Leveraging sentiment analysis at the aspects level to predict ratings of reviews. Inf. Sci. 451–452, 295–309. https://doi.org/10.1016/j.ins.2018.04.009.
Ramanathan, R., 2011. An empirical analysis on the influence of risk on relationships between handling of product returns and customer loyalty in E-commerce. Int. J. Prod. Econ. 130, 255–261. https://doi.org/10.1016/j.ijpe.2011.01.005.
Rao, S., Lee, K.B., Connelly, B., Iyengar, D., 2018. Return time leniency in online retail: a signaling theory perspective on buying outcomes. Decis. Sci. J. 49, 275–305. https://doi.org/10.1111/deci.12275.
Rejikumar, G., Asokan-Ajitha, A., Dinesh, S., Jose, A., 2022. The role of cognitive complexity and risk aversion in online herd behavior. Electron. Commer. Res. 22. https://doi.org/10.1007/s10660-020-09451-y.
Robertson, T.S., Hamilton, R., Jap, S.D., 2020. Many (Un)happy returns? The changing nature of retail product returns and future research directions. J. Retailing 96, 172–177. https://doi.org/10.1016/j.jretai.2020.04.001.
Röder, M., Both, A., Hinneburg, A., 2015. Exploring the space of topic coherence measures. In: WSDM 2015 - Proceedings of the 8th ACM International Conference on Web Search and Data Mining, pp. 399–408. https://doi.org/10.1145/2684822.2685324.
Rokonuzzaman, M., Iyer, P., Harun, A., 2021. Return policy, No joke: an investigation into the impact of a retailer's return policy on consumers' decision making. J. Retailing Consum. Serv. 59, 102346. https://doi.org/10.1016/j.jretconser.2020.102346.
Sahoo, N., Dellarocas, C., Srinivasan, S., 2018. The impact of online product reviews on product returns. Inf. Syst. Res. 29, 723–738. https://doi.org/10.1287/isre.2017.0736.
Shang, G., Ghosh, B.P., Galbreth, M.R., 2017a. Optimal retail return policies with wardrobing. Prod. Oper. Manag. 26, 1315–1332. https://doi.org/10.1111/poms.12690.
Shang, G., Pekgün, P., Ferguson, M., Galbreth, M., 2017b. How much do online consumers really value free product returns? Evidence from eBay. J. Oper. Manag. 53–56. https://doi.org/10.1016/j.jom.2017.07.001.
Siering, M., Deokar, A.V., Janze, C., 2018. Disentangling consumer recommendations: explaining and predicting airline recommendations based on online reviews. Decis. Support Syst. 107, 52–63. https://doi.org/10.1016/J.DSS.2018.01.002.
Sluis, S., De Giovanni, P., 2016. The selection of contracts in supply chains: an empirical analysis. J. Oper. Manag. 41, 1–11. https://doi.org/10.1016/j.jom.2015.10.002.
Tirunillai, S., Tellis, G.J., 2014. Mining marketing meaning from online chatter: strategic brand analysis of big data using latent dirichlet allocation. J. Mark. Res. 51, 463–479. https://doi.org/10.1509/JMR.12.0106/ASSET/IMAGES/LARGE/10.1509_JMR.12.0106-FIG8.JPEG.
Tonning, W.A., 1956. The beginnings of the money-back guarantee and the one-price policy in champaign-urbana, Illinois, 1833–1880. Bus. Hist. Rev. 30. https://doi.org/10.2307/3111973.
Ülkü, M.A., Dailey, L.C., Yayla-Küllü, H.M., 2013. Serving fraudulent consumers? The impact of return policies on retailer's profitability. Serv. Sci. 5. https://doi.org/10.1287/serv.2013.0051.
Van Nguyen, T., Zhou, L., Chong, A.Y.L., Li, B., Pu, X., 2020. Predicting customer demand for remanufactured products: a data-mining approach. Eur. J. Oper. Res. 281, 543–558. https://doi.org/10.1016/j.ejor.2019.08.015.
Walmart, 2025. Guides | seller performance standards. https://marketplacelearn.walmart.com/guides/Policies%20%26%20standards/Performance/Seller-performance-standards.
Walsh, G., Albrecht, A.K., Kunz, W., Hofacker, C.F., 2016. Relationship between online retailers' reputation and product returns. Br. J. Manag. 27, 3–20. https://doi.org/10.1111/1467-8551.12120.
Wang, B., Wang, P., Tu, Y., 2021. Customer satisfaction service match and service quality-based blockchain cloud manufacturing. Int. J. Prod. Econ. 240. https://doi.org/10.1016/j.ijpe.2021.108220.
Woimant, A., Steils, N., 2025. "Take no risk!": immediate and delayed backfire-effect of risk-reducing messages in retailers' return and exchange policies. J. Retailing Consum. Serv. 82, 104134. https://doi.org/10.1016/j.jretconser.2024.104134.
Wood, S.L., 2001. Remote purchase environments: the influence of return policy leniency on two-stage decision processes. J. Mark. Res. 38, 157–169. https://doi.org/10.1509/jmkr.38.2.157.18847.
Xu, L., Li, Y., Govindan, K., Xu, X., 2015. Consumer returns policies with endogenous deadline and supply chain coordination. Eur. J. Oper. Res. 242, 88–99. https://doi.org/10.1016/j.ejor.2014.09.049.
Xu, X., Lee, C., 2020. Utilizing the platform economy effect through EWOM: does the platform matter? Int. J. Prod. Econ. 227, 107663. https://doi.org/10.1016/j.ijpe.2020.107663.
Yan, R., Cao, Z., 2017. Product returns, asymmetric information, and firm performance. Int. J. Prod. Econ. 185, 211–222. https://doi.org/10.1016/j.ijpe.2017.01.001.
Yan, T., Wieland, A., Tate, W., 2025. Researching like a master chef: an expansion of the quantitative "kitchen tools" in supply chain management research. J. Supply Chain Manag. 1–10. https://doi.org/10.1111/JSCM.12347, 0.
Zhang, D., Frei, R., Senyo, P.K., Bayer, S., Gerding, E., Wills, G., Beck, A., 2023. Understanding fraudulent returns and mitigation strategies in multichannel retailing. J. Retailing Consum. Serv. 70, 103145. https://doi.org/10.1016/J.JRETCONSER.2022.103145.
Zhang, J., Li, H., Yan, R., Johnston, C., 2017. Examining the signaling effect of e-tailers' return policies. J. Comput. Inf. Syst. 57. https://doi.org/10.1080/08874417.2016.1183989.
Zhao, Y., Xu, X., Wang, M., 2019. Predicting overall customer satisfaction: big data evidence from hotel online textual reviews. Int. J. Hospit. Manag. 76. https://doi.org/10.1016/j.ijhm.2018.03.017.