Impact of delivery performance on online review ratings: the role of temporal distance of ratings

Prashanth Ravula

Revised: 27 July 2021 / Accepted: 23 April 2022 / Published online: 18 May 2022
© The Author(s), under exclusive licence to Springer Nature Limited 2022

Abstract
Customers are increasingly using online reviews in their purchase decision-making processes. As sellers benefit from displaying several reviews with favorable ratings, many sellers solicit reviews from customers. When a customer places an order on an e-commerce platform, the seller gets a notification to fulfill the order, and the customer is notified of the estimated delivery date. Some customers receive their products on time, while others receive their orders either earlier or later than the notified delivery date. After customers receive their products, the sellers often solicit reviews. This research focuses on the impact of delivery performance on review ratings. Specifically, this study addresses two questions: (1) Do customers reward sellers for early delivery in the same way they penalize them for late deliveries? (2) What is the role of the temporal distance of rating in online ratings in the context of delivery performance? The study estimates ordinal logit models in the Bayesian framework. Findings of the study indicate that customers give much lower (a little higher) ratings to orders delivered late (early) than to orders delivered on time. Further, the findings indicate that temporal distance is positively associated with ratings for late deliveries. The study discusses the theoretical and managerial implications of these results.

Keywords Online reviews · e-commerce · Delivery performance · Bayesian models

Introduction
With the increased diffusion of smartphones and the internet, e-commerce retail sales have seen significant growth in the past decade (Statista 2021). Given this phenomenon, many retailers have been adding online channels to reach more customers and increase sales revenue. Further, e-commerce platforms have created new opportunities for resource-constrained small retailers, allowing them to sell merchandise through the platforms. On the other hand, the number of consumers shopping online has also seen tremendous growth. According to an industry estimate, over two billion consumers have purchased goods or services online in 2020 (Statista 2022).

The recent global coronavirus (Covid-19) pandemic has led to a surge in e-commerce and accelerated these trends. A recent report shows the strong uptake of e-commerce sales across the world, with many consumers, especially in emerging economies, making the greatest shift to online shopping (UNCTAD 2022). For example, a study by AppsFlyer (2021) reveals that the number of users who downloaded an e-commerce app on smartphones has increased by 48% in 2021. In addition, many consumers expanded the type of products they purchase online. As consumers are getting used to the convenience of online shopping, they are more likely to stay with that habit of online shopping (SBT 2022).

Perhaps, reflecting the above trends, industry estimates (eMarketer 2022) indicate that e-commerce retail sales are expected to increase to $7.39 trillion by 2025, accounting for 23.6% of global retail sales. This growth in e-commerce retail sales and competition among sellers require sellers to provide fast and convenient delivery, as after-sales service can impact customer satisfaction. As such several sellers collect customer satisfaction ratings after orders were delivered.

Online reviews have increasingly become an alternative to traditional customer satisfaction surveys. When a customer places an order on an e-commerce platform (e.g., Alibaba, Amazon, Flipkart, Olist Store), the seller gets a notification to fulfill the order, and the customer is notified of the estimated delivery date. Some customers receive products on...
time, while other customers receive their orders either earlier or later than the notified delivery date. After customers receive their products, the sellers often solicit reviews from customers. Specifically, as shown in Fig. 1, sellers ask customers to rate the experience with the seller. While customers tend to comprehend purchase experience with the seller on an e-commerce platform in a holistic perspective, it can be assumed that delivery performance will affect the overall customer satisfaction reflected in rating because these reviews are solicited right after delivery of orders. As such, we use online review ratings to evaluate the relationship between delivery performance and customer satisfaction.

Further extant research suggests that prospective customers increasingly use online reviews in their purchase decisions (Changchit and Klaus 2020; Berger et al. 2010; Chevalier and Mayzlin 2006; Zhu and Zhang 2010). For example, a recent survey (Carter 2022) found that around 89% of consumers agree that online reviews are an essential part of the purchase process. However, despite the growing importance of online reviews in e-commerce retail, surprisingly little is known about the relative effects of delivery performance: early and late vs. same-day delivery on online review ratings. On the other hand, some customers post their reviews, rating the sellers, on the same day of delivery, and some do it later. Extant research suggests that temporal distance\(^1\) of rating influences online review ratings (Huang et al. 2016). However, there is no published research on the role of the temporal distance of ratings in the context of e-commerce delivery performance to the best of our knowledge. Filling the above gaps in the literature, this study, specifically, investigates two key questions in the context of after-sales service in the e-commerce retail sector: (1) Do customers reward sellers for early delivery in the same way they penalize them for late deliveries? (2) What is the role of the temporal distance in online review ratings in the context of delivery performance?

For the empirical analysis, e-commerce platform data were used from an emerging market. Data consist of late, early, and same-day deliveries. However, these incidences were not randomized. The possibility also existed that early and late deliveries were systematically different from same-day deliveries. A propensity score matching (PSM) algorithm was used to generate matched samples to address these issues. We then estimated ordinal logistic regression models on the matched samples. Specifically, two models were estimated: early vs. same-day deliveries and late vs. same-day deliveries. Insights from this study are managerially relevant. As e-commerce retail growth has become more prevalent in emerging markets than in developed countries (Kuhn and Petzer 2018), these findings assist retailers in devising effective review solicitation strategies. As such, this research represents the first study that underscores the importance of review solicitation time in the context of late deliveries, contributing to different streams of literatures: prospect theory, construal level theory, online reviews, and retail.

Theoretical background

As the e-commerce sector of retailing is rapidly growing, many firms are transforming their distribution channels to reengineer their relationships with customers. Several new features or drivers of customer satisfaction were identified in this online setting. The first set of studies focused on features related to the internet, such as ease of use, trust, etc. (Bhatnagar et al. 2000; Zeithaml et al. 2002). As purchase experience with e-commerce firms can be evaluated according to a range of features, the second set of studies identified several other features related to personalization, product range, prices, checkout, shipping, etc. as drivers of customer satisfaction (Dholakia and Zhao 2010; Jin and Park 2006). While several studies focused on the pre-delivery process, Jiang and Rosenbloom (2005) investigated and showed that customer satisfaction can vary between e-commerce checkout and after delivery, indicating delivery performance is a

---

\(^1\) Temporal distance is measured as the difference between the actual delivery date and the review posting date in the number of days.
critical touchpoint in consumers’ overall satisfaction with the
sellers. Later research provided ample evidence, recog-
nizing delivery performance (such as online time delivery, total delivery time) as a significant determinant of overall
customer satisfaction in the e-commerce sector of retailing
(Blut 2016; Collier and Bienstock 2006; Dholakia and Zhao
2010; Jain et al. 2015; Thirumalai and Sinha 2005; Vaku-
lenko et al. 2019).

Traditionally, practitioners and researchers used surveys
to understand the impact of delivery performance on cus-
tomer satisfaction. However, as an alternative to customer
satisfaction surveys, online reviews have become more
popular during the past two decades (Rese et al. 2014). As
discussed earlier, after delivering the orders, e-commerce
platforms ask customers to rate their experience with sellers.
As these reviews are solicited after delivery of the order, it is
reasonable to expect that review ratings may reflect customer
satisfaction due to delivery performance. As such, for evalu-
ating the relationship between delivery performance and
customer satisfaction, we use online review ratings proxy
for customer satisfaction.

Further, extant research suggests that prospective con-
sumers are increasingly relying on these online reviews for
their purchase decisions (Changchit and Klaus 2020). Per-
haps reflecting this, online reviews, especially the review
ratings (Luca 2011), have impacted sales (e.g., Berger et al.
2010; Chevalier and Mayzlin 2006; Zhu and Zhang 2010).
As such, several studies focused on determinants of online
review ratings. However, most of these studies in online
reviews literature were focused on reviews related to con-
sumption experience with products.

A few studies have investigated the drivers behind review
ratings in the context of e-commerce delivery. Qu et al.
(2008) investigated the driving factors behind the review
ratings on Yahoo! Merchants and found that the review
ratings increased with on-time delivery. Park et al. (2012)
examined the differential effects of pre- and post-transaction
performance such as the fulfilled delivery, order tracking,
and customer support on online review ratings using data
from BizRate.com. Wu et al. (2021) focused on the relative
effects of coupons vs. free shipping on review ratings for an
e-commerce website. Li and Wang (2021), using reviews
from Amazon, found that more retailers using free shipping
increased product review ratings.

A few other studies focused on temporal distance in dif-
ferent contexts on review ratings (Huang et al. 2016; Li
et al. 2019; Stamolampros and Korfiatis 2018; Yang et al.
2018; Wu et al. 2021). For example, in the context of restau-
rants, Huang et al. (2016) found that the temporal distance
of review and consumption has a positive effect on review
rating. Wu et al. (2021), in the context of e-commerce pur-
cases, found that coupons increase review ratings through
perceptions of monetary savings when the temporal distance
of purchase and review is close but decrease review ratings
through low perceived product quality when temporal dis-
tance is far.

The above literature suggests that factors related to
e-commerce delivery and temporal distance of ratings have
a role in review ratings. However, there are a few insights
into questions like (1) Do customers reward sellers for early
delivery similarly as they penalize late deliveries? (2) What
is the role of the temporal distance in online review ratings
in the context of delivery performance? These are the gaps
that we address in this research.

Conceptual framework

This section will discuss the conceptual framework used for
this study and propose corresponding hypotheses.

Consumer evaluation of early and late vs. same-day
deliveries

For e-commerce orders, consumers expect to be notified
when they receive their orders. As discussed earlier, sellers
notify consumers with promised delivery dates after process-
ing orders. In some cases, sellers deliver the orders earlier
than notified delivery date. Some consumers receive on the
same day of the notified delivery date, while some other con-
sumers receive later. Against this backdrop, drawing from
prospect theory, we argue that delayed deliveries are experi-
enced with greater psychological force than early deliveries
of similar magnitude, affecting review ratings (Chan et al.
2018; Gal and Rucker 2018; Kahneman and Tversky 2013;
Thaler 2000).

Prospect theory (PT), particularly, asserts that consum-
ers are more attuned to differences (relative to a reference
point) and inclined to place greater weights on losses than
gains of an equal magnitude (Kahneman and Tversky 2013;
Tversky and Kahneman 1992). In these lines, Thaler (2000,
p.137) argued that “losses hurt about twice as much as gains
make us feel good.” Extant research applied prospect theory
to several different contexts. For example, PT is applied to
the sustainable operation of transport infrastructure pro-
jects under government regulation (Ma et al. 2021), to pilot
weather-related decision-making in an uncertain situation
involving monetary gains and risk-seeking (Walmsley and
Gilbey 2020), to explain investment strategies under uncer-
tainly (Frazzini 2006), to product pricing strategy (Koszegi
and Rabin 2006), to explain the effectiveness of promotional
prices of leisure services (Crompton 2016), and to travel-
ers’ behavior in situations involving travel time uncertainty
(Ramos et al. 2014). Similarly, asymmetric disconfirmation
in satisfaction literature suggests that consumers’ negative
consumption experiences have a greater influence on their
judgment than positive experiences (Anderson and Sullivan
A few studies focused on online review ratings have confirmed this phenomenon (Chan et al. 2018; Hao et al. 2010; Moe et al. 2011). The above literature suggests a steeper value function in the loss region as compared to gains. In other words, individuals evaluate potential losses differently from potential gains with equal magnitude.

Extending the above discussion to the present context of e-commerce retail transactions related to online reviews, this study argues that consumers evaluate early and late deliveries differently; specifically, that consumers use same-day delivery as a reference and compare early and late deliveries with the same magnitude (e.g., a day late/early), giving more weight to the late deliveries than the early deliveries. Based on the above information, it is predicted that consumers will penalize sellers more for late deliveries (e.g., one day late) than reward sellers for early deliveries with similar magnitudes (e.g., one day early). Therefore, the following hypothesis is proposed:

**H1** The negative effect of a late delivery on a review rating will be larger than the positive effect from early delivery.

### Role of temporal distance

The temporal distance can be defined as the time between two events. In the present context, temporal distance refers to the difference (in days) between an order delivery date and the review posting date. Extant literature shows that temporal distance is one of the key psychological distances that shape consumers’ judgments (Adler and Sarstedt 2021; Mishra et al. 2020; Trope and Liberman 2010). Specifically, the construal level theory (CLT) suggests that consumers’ memories related to events (e.g., consumption experience) are inconsistent with their perception of those events at the time they happen (Trope and Liberman 2010). When consumers recall their experiences with those events for making judgments, they tend to use different mental representations (construals) depending on their perceived psychological distance from those events (Trope and Liberman 2010; Yuddkin et al. 2020; Wu et al. 2021). For instance, if consumers perceive the greater psychological distance between themselves and those events, they process events at higher levels of construal, abstractly. In contrast, if consumers perceive less psychological distance, they process the event at lower levels of construal, thoroughly or concretely. The above discussion suggests that events occurring at relatively closer temporal proximities are processed differently than events occurring in more distant proximities. Such differences in the processing of events are expected to influence consumer evaluations of those events (Adler and Sarstedt 2021; Mishra et al. 2020; Trope and Liberman 2010).

A recent study (Wang and Lin 2021) investigated the effect of temporal distance on consumer price evaluation. Specifically, the authors show that when the temporal distance is near, a nine-ending price may be perceived as larger than a price that is actually one dollar higher. However, the perceived magnitude of difference due to the left-digit effect has diminished when the temporal distance is distant. Another study (Liu et al. 2020), focused on temporal distance in consumer evaluation of online promotion activities and purchase behavior, found that temporal distance has a positive (negative) impact on purchase decision of high (low) involvement products. Choi et al. (2019) study found that consumers perceive partitioned pricing as more attractive than combined pricing for a temporally distant event. Su et al. (2022) found that consumers evaluate travel items into more superordinate (subordinate) categories when it comes to distance-future (near-future) trips. The above literature suggests that consumers, while evaluating different events/objects/decisions, use different mental representations (abstract vs. concrete) depending on the temporal distance.

A few studies, more relevant to the present study, have focused on the effect of temporal distance on review ratings given to consumption experiences (Huang et al. 2016; Li et al. 2019; Stamolampros and Korfiatis 2018; Yang et al. 2018; Wu et al. 2021). All these studies found that temporal distance to consumption experience is positively associated with review rating. The above literature suggests that temporal distance influences consumer evaluations.

In the context of the e-commerce sector of retailing transactions related to online reviews, it is expected that temporal distance (psychological distance) will affect how consumers construe their purchase experiences with sellers. Specifically, on the same day of delivery, the event (i.e., delivery performance) is psychologically very close and can be processed in a detailed, concrete manner. As the days pass, however, the event will be construed more abstractly, and detailed aspects of the delivery performance will gradually fade (Kim et al. 2008). In such situations, consumers are more likely to rely on their overall experiences with the sellers. Research also suggests that positive aspects of an event are more salient when processing abstractly, and pros are easier to think of when considering temporally distance versus close events (William et al. 2014; Huang et al. 2016). Therefore, the more time between the delivery date and the date on which they post their reviews, the greater the temporal distance. In this instance, consumers are more likely to make more positive review ratings. Thus, the following hypothesis is proposed:

**H2** The temporal distance between the delivery date and the review posting date is positively associated with the review rating.
Impact of delivery performance on online review ratings: the role of temporal distance of ratings

Following literature, we have included several control variables for accounting observed heterogeneity: shipping fee (Kim and Cheon 2020; Kulkarni 2020; Ma 2017; Wu et al. 2021), spatial distance (Arentze and Timmermans 2001; Blut et al. 2018); product category (Kim 2020; Kim and Cheon 2020); and season (Lee et al. 2018). We present the conceptual framework in Fig. 2.

Methodology

Data

In order to test these hypotheses, data from e-commerce company in an emerging market were used for this study. After a customer purchases a product from the e-commerce store, the seller gets notified to fulfill that order. After the seller ships the product, customer is notified of the delivery date. Once the customer receives the product, sellers solicit reviews. The dataset primarily consists of order information, customer zip code, seller zip code, and respective online review related information. For this study, only orders with a single product from a single seller were used.

Variables

rating<sub>io</sub> is an ordinal variable used to indicate the customer evaluation of his/her purchase experience with the seller. Specifically, a review rating refers to the number of stars (1 to 5) allocated by customer <i>i</i> when indicating his/her assessment of his/her purchase experience with the seller for order <i>o</i>.

<sub>del_days</sub><sup>io</sup> is a continuous variable that represents the relative deliver days measured as the deviation of the actual delivery date from the notified delivery date. For example, if a retailer promised a two-day delivery and the package arrived after (within) one day, then the <sub>del_days</sub><sup>io</sup> would be equal to one day late (one day early).

<sub>temp_dist</sub><sup>io</sup> is a continuous variable that represents the temporal distance of the review rating by customer <i>i</i> for order <i)o</i>. It was measured as the difference between the actual delivery date and the review posting date in number of days.

<sub>pcat_conv</sub><sup>io</sup> is an indicator variable that represents whether the product purchased was from a convenience product category. Following the literature (Nguyen et al. 2019; Thirumalai and Sinha 2005), products were classified into three categories: convenience goods, shopping goods, and specialty goods. This classification is based on the volume and unit value of the products purchased. For example, consumers tend to buy convenience goods in large volumes and at low unit costs. <sub>pcat_conv</sub><sup>io</sup> was set to 1 if the product purchased by customer <i>i</i> in order <i>o</i> was from a convenience product category, 0 otherwise.

<sub>pcat_splt</sub><sup>io</sup> is an indicator variable that represents whether the product purchased was from a specialty product category. Similar to the convenience product category, <sub>pcat_splt</sub><sup>io</sup> was set to 1 if the product purchased by customer <i>i</i> in order <i>o</i> was from a specialty product category, 0 otherwise.

<sub>spat_dist</sub><sup>io</sup> is a continuous variable indicating the distance between the seller’s city and customer’s city. Using the latitude and longitude of the city zip codes, the spatial distance between the seller and customer was computed in kilometers. For example, if the customer and seller are from the same city (i.e., same zip codes), then the spatial distance between the seller’s city and customer’s city was equal to zero.

<sub>shipfee</sub><sup>io</sup> is a continuous variable measured relative to order value. For example, if the shipping costs were $30 for an order worth $30, then the respective <sub>shipfee</sub><sup>io</sup> was 1. Similarly, if shipping costs were $30 for an order worth $100, then the respective <sub>shipfee</sub><sup>io</sup> was 0.3.

<sub>season</sub><sup>io</sup> is an indicator variable that measured whether the purchase was completed during a holiday season (i.e., November and December). Therefore, if customer <i>i</i> placed order <i>o</i> in either November or December, then <sub>season</sub><sup>io</sup> was set to 1, 0 otherwise.

Model

The dependent variable, review rating, is an ordinal variable consisting of five ordered categories. Therefore, the Ordinal Logit Model, the best-fitting statistical model for handling an ordered outcome (McCullagh 1980), was used to investigate the effect of delivery performance on a customer’s review rating. Assuming that the utility of customer <i>i</i> is represented by an unobservable latent variable <i>U</i><sub>i</sub>, then customer <i>i</i> gives a certain rating <sub>rating</sub><sup>io</sup> between 1 and 5 on the basis of <i>U</i><sub>i</sub>.

<math>
U_i = x_i'\beta + \epsilon_i 
</math>

(1)

<math>
\text{Rating}_{i} = j \text{ if } \theta_{j-1} < U_i \leq \theta_j, 
</math>

(2)
Table 1  Descriptive statistics

| Variable       | Rating        | Early delivery | Same-day delivery | Late delivery | Full sample |
|----------------|---------------|----------------|-------------------|---------------|-------------|
| temporal distance | distio        | 488.956        | 507.113           | 471.895       | 489.321     |
| relative delivery days | del_daysio    | 11.175         | 0                 | 1.920         | NA          |
| shipping fee    | shipfeeio     | 0.260          | 0.263             | 0.267         | 0.263       |
| convenience goods | pcat_convio   | 0.236          | 0.290             | 0.271         | 0.265       |
| shopping fee    | pcat_shopio   | 0.650          | 0.586             | 0.576         | 0.604       |
| specialty goods | pcat_spltio   | 0.115          | 0.124             | 0.083         | 0.131       |
| season          | seasonio      | 0.070          | 0.064             | 0.083         | 0.072       |
| number of observations | 314           | 314            | 314               | 942           |

Matched sample

The challenge in using secondary data is that the incidences in the same-day delivery, early delivery, and late delivery groups are not randomly selected. The matched sample approach essentially attempts to address this issue by creating a pseudo-random sample. In recent years, the propensity score matching approach has gained popularity because it allows a refined matching process along multiple characteristics (Dehejia and Wahba 1999). In essence, this approach attempts to correct for the non-random treatment effect by matching a treated incidence (early or late delivery) to an untreated incidence (same-day delivery) that has similar observed characteristics. Although the results from the matched samples do not establish a causal relationship, they provide evidence that the observed relationship is related to the delivery performance and temporal distance rather than other observed characteristics.

The R package TriMatch (Bryer 2013) was used for this study as it estimates propensity scores and finds the best matched triplets with replacement. Shipping fee, temporal distance, product type, and season were used for the selection of the matched samples. The details for the propensity score matching procedure are provided in Appendix. The resulting matched sample consisted of 942 incidences (314 per each group).

Model estimation

For ease of comparison, two models were estimated. First, a model for early deliveries vs. same-day deliveries was estimated. In other words, a model (Eqs. 1–4) was estimated for the data consisting of 314 early deliveries and 314 matched same-day deliveries. Similarly, another model was estimated for late deliveries vs. same-day deliveries. For this model, data consisting of 628 observations (314 late deliveries and 314 matched same-day deliveries) were utilized.

For inference regarding the parameters, the PROC MCMC method in SAS with highly diffuse priors for the model parameters was used. Specifically, the Gaussian prior distribution for all parameters in $\beta$ and $\theta$ was used. For each model (early deliveries and late deliveries), an MCMC chain with 50,000 samples was simulated and the first 10,000 samples were discarded as burn-in. From the remaining samples, every 10th iteration was selected, allowing for a retention of 4000 samples for posterior inference of means and standard deviations of the parameter estimates.

Results

Descriptive statistics and correlations are provided in Tables 1 and 2. For the early delivery sample, the average review rating was 4.411. While same-day deliveries received a lower average rating (4.156) than early deliveries, late deliveries had the lowest average rating (3.452). One of the key variables was relative delivery days (i.e., the number of days before (after) the notified delivery date for early (late) deliveries). For same-day deliveries, the relative delivery days was zero, which was the reference point for consumers to compare their deliveries against. Early deliveries were delivered, on average, 11.175 days before this date and late deliveries were delivered, on average, 1.920 days after the notified delivery date.

The most interesting insight from this analysis was related to the temporal distance of the review rating. Consumers
Impact of delivery performance on online review ratings: the role of temporal distance of ratings

who received early deliveries posted their reviews, on average, a day (1.000) after they received their deliveries. However, consumers who received late deliveries posted their reviews, on average, 0.678 days after they received their deliveries (i.e., less than a day). Four variables (i.e., shipping fee, spatial distance, type of good, season) were used to select the matched samples (see Appendix). Reflecting on this information, perhaps the matched samples had similar descriptive statistics for these four variables across the three samples.

The correlations between different variables are presented in Table 2. An interesting correlation was between the review rating and temporal distance. As temporal distance increased, so did the rating.

### Results for the early deliveries model

To estimate the effect of early deliveries on review ratings, Eqs. 1–4 were utilized for the sample of 314 early-day deliveries and 314 same-day deliveries. The parameter estimates for the early deliveries model are displayed in Table 3. Consistent with this study’s expectations, the coefficients of $rdd_{io}$ had the expected sign and significance (i.e., 95% posterior distribution of the difference of means excluding zero). The coefficient of $rdd_{io}$ was statistically significant and positive (0.119, sig. = 0.05), implying that early deliveries were positively associated with the review ratings. The coefficient of $temo_{dist_{io}}$ was positive, but not statistically significant (0.055).

### Results for the late deliveries model

To estimate the effect of late deliveries on review ratings, Eqs. 1–4 were utilized for the sample of 314 late-day deliveries and 314 same-day deliveries. The parameter estimates for the late deliveries model are displayed in Table 4. Consistent with this study’s expectations, the coefficients of $rdd_{io}$ and $temo_{dist_{io}}$ had the expected signs and significance (i.e., 95% posterior distribution of the difference of means excluding zero). The coefficient of $rdd_{io}$ was statistically significant and negative ($-0.610$, sig. = 0.05), implying that late deliveries were negatively associated with the review ratings. In other words, the longer the delivery was past the expected date, the lower the review rating. The coefficient of $temo_{dist_{io}}$ was positive and statistically significant (1.109, sig. = 0.05), implying that the temporal distance of the review rating was positively associated with the review rating.
The above results supported this study’s hypotheses. First, the comparison of the parameters for relative delivery days in the early (0.119, sig. = 0.05) and late delivery models (−0.610, sig. = 0.05) indicated that the negative effect of late deliveries on consumers’ review ratings was larger than the positive effect of early deliveries with similar magnitudes. This result was consistent with the prediction from the Prospect theory, which suggests that the effect from the perceived loss was stronger than the effect from the perceived gains (Chan et al. 2018; Kahneman and Tversky 2013). This finding supported the H1.

Second, the parameters for temporal distance in early (0.055) and late delivery models (1.109 sig. = 0.05) indicated that the temporal distance of the review rating was positively associated with the review rating. This result was consistent with the predictions from the Construal Level theory (Huang et al. 2016; Trope and Liberman 2010; Yudkin et al. 2020; Wu et al. 2021), which indicates that consumers with the greater distance between the event and themselves, more favorable rating. However, the effect of temporal distance for early deliveries was not significant. These findings support H2 partially.

Conclusions

Consumers are increasingly using online reviews in their purchase decisions. As such, retailers are soliciting reviews from their customers after delivering the products. In the context of e-commerce delivery performance, this study aimed to answer two key research questions: (i) do customers reward sellers for early delivery in the same way they penalize them for late deliveries? and (ii) what is the role of the temporal distance in online review ratings in the context of delivery performance? E-commerce data from a particular company were used within an emerging economy for this study. To establish the causal relationship, a propensity score-matched sample was used. Two models: early deliveries and late deliveries were estimated. The empirical results supported the proposed hypotheses. Specifically, we hypothesized (in H1) that the negative effect of a late delivery on a review rating will be larger than the positive effect from early delivery. The study findings indicated that customers gave much lower (a little higher) ratings to orders delivered late (early) than to orders delivered on time. These results supported H1 and answered the first research question of the study. Further, we hypothesized (in H2) that the temporal distance between the delivery date and the review posting date is positively associated with the review rating. Consistent with our prediction, the study’s findings indicated that temporal distance was positively associated with review ratings. This result supported H2 and answered the second research question of the study.

Overall, the main contributions of this study to the existing body of literature are threefold. Specifically, the research highlights the value of investigating relative delivery days and the temporal distance of review ratings. It also contributes to the literature on prospect theory, construal level theory, and online reviews.

First, the findings extend an array of prior studies using prospect theory (Chan et al. 2018; Kahneman and Tversky 2013) and indicated that customers give much lower (a little higher) ratings to orders delivered late (early) than to orders delivered on time, consistent with the predictions of loss aversion from prospect theory. Second, the findings extend an array of prior studies using Construal Level theory (e.g., Huang et al. 2016; Trope and Liberman, 2010), which has been researched in wide range of disciplines, such as marketing, organizational study, psychology, and education. Third, this study extended the online review literature by applying prospect theory and construal level theory in a context of significant relevance to practitioners and academics in a growing sector of retail: e-commerce. The majority of prior studies on online reviews has focused on understanding the consequences of review ratings, such as sales (Chevalier and Mayzlin 2006; Zhu and Zhang 2010). This study, in contrast, examined the antecedents of online review ratings (e.g., Chen and Kirmani 2015; Huang et al. 2016).

Managerial implications

E-commerce delivery performance is viewed as one of the core features of customer satisfaction (Vakulenko et al. 2019) reflected in online reviews. As prospective consumers increasingly use reviews in their purchase decisions (Chevalier and Mayzlin 2006), sellers would benefit from having more favorable ratings. As such, sellers solicit reviews after delivery of orders. Lemon and Verhoef (2016) suggest sellers can influence customer satisfaction at touchpoints owned by sellers, such as providing a seamless shopping experience at e-commerce platforms, obtaining favorable ratings. However, customer satisfaction in the case of late deliveries is beyond sellers’ control; therefore, sellers require some degree of management and a strategy for soliciting ratings. The study’s findings will help firms devise when to solicit online reviews in cases of late deliveries.

In general, sellers on e-commerce sites solicit online reviews immediately after a shipment has been delivered (i.e., on the same day as the delivery). The findings of this study indicated that consumers who received late deliveries posted their ratings, on average, on the same day and gave much lower ratings. Therefore, the findings indicated that the temporal distance of the review ratings was positively associated with the review ratings. In other words, if customers post their reviews a day after receiving their deliveries, they may give a higher rating.
Therefore, sellers may avoid getting penalized for late deliveries by soliciting online reviews later rather soliciting online reviews on the same day of delivery. Second, the findings suggested that retailers do not gain much benefit in terms of favorable ratings for early deliveries. Therefore, sellers can avoid promising longer delivery dates and delivering early.

Limitations and future research

Although this study has provided a better understanding of the effects of delivery performance and temporal distance on review ratings in the context of e-commerce, some limitations exist that could be addressed in future studies. First, this study used secondary data. Although a propensity score-matched sample was employed to investigate causality, the data are still considered weak in regard to discovering the underlying reasons for the phenomenon. Future studies should employ experimental methods that provide more detailed reasoning about the relationships identified in the study. Second, other issues related to delivery performance, such as the quality of the shipment (e.g., damages to the product during shipment) and the customer’s relationship duration with the seller, could be added to the model in future studies to achieve greater explanatory power in regard to consumer satisfaction ratings. Third, as this study used data from an emerging economy, generalizations related to the findings should be made with caution. In the future, other researchers should conduct cross-country comparison studies so that the results can be more easily generalized.

Appendix: propensity score matching

For selecting a matched sample, we use R package TriMatch, which estimates the propensity scores and finds best matching triplets. We employ different variables as matching variables: Shipping fee, Spatial distance, Type of product (e.g., convenience, shopping, specialty, and season). Further, we compute the Standardized Bias (SB), a widely used technique to ensure the balance of the samples (Harmeling et al. 2015). A SB score below 0.1 indicates the PSM is effective in balancing the distributions of the covariates. We report the SB Scores in Table 5.

\[
SB_{\text{match}} = \frac{M_1(X_k) - M_0(X_k)}{\sqrt{0.5(V_1(X_k) - V_0(X_k))}}. \tag{5}
\]

where \(M_1(X_k)\) and \(M_0(X_k)\) are the means [variances] of the observable \(k\) for the treated group and the matched control group.

Table 5  Standardized bias scores for matched sample

| Variable          | Early delivery to be more Same-day delivery | Late delivery to be more Same-day delivery |
|-------------------|---------------------------------------------|-------------------------------------------|
|                   | Standardized bias                           | Standardized bias                         |
| Shipping fee      | -0.01                                       | 0.02                                      |
| Log (spatial distance) | -0.04                                       | -0.08                                     |
| Convenience goods | -0.12                                       | -0.04                                     |
| Specialty goods   | -0.03                                       | 0.08                                      |
| Season            | 0.03                                        | 0.07                                      |

References

Adler, S., and M. Sarstedt. 2021. Mapping the jungle: A bibliometric analysis of research into construal level theory. Psychology & Marketing 38 (9): 1367–1383.

Anderson, E.W., and M.W. Sullivan. 1993. The antecedents and consequences of customer satisfaction for firms. Marketing Science 12 (2): 125–143.

Appsflier. 2021. The state of ecommerce app marketing. Retrieved April 10, 2022, https://www.appsflier.com/infographics/state-of-ecommerce-2021/

Arentze, T.A., and H.J. Timmermans. 2001. Deriving performance indicators from models of multipurpose shopping behavior. Journal of Retailing and Consumer Services 8 (6): 325–334.

Berger, J., A.T. Sorensen, and S.J. Rasmussen. 2010. Positive effects of negative publicity: When negative reviews increase sales. Marketing Science 29 (5): 815–827.

Bhatnagar, A., S. Misra, and H.R. Rao. 2000. On risk, convenience, and internet shopping behavior. Communications of the ACM 43 (11): 98–105.

Blut, M. 2016. E-service quality: Development of a hierarchical model. Journal of Retailing 92 (4): 500–517.

Blut, M., C. Tellor, and A. Floh. 2018. Testing retail marketing-mix effects on patronage: A meta-analysis. Journal of Retailing 94 (2): 113–135.

Bryer, J.M. 2013, May. TriMatch: An R package for propensity score matching of non-binary treatments. In The R user conference, useR (pp. 10–12).

Carter, R. 2022. The ultimate list of online review statistics for 2022. https://findstack.com/online-review-statistics/. Accessed March 04, 2022.

Chan, T., Z. Liu, & W. Zhang. 2018. Delivery service, customer satisfaction and repurchase: Evidence from an online retail platform. Customer satisfaction and repurchase: Evidence from an online retail platform (September 15, 2018).

Changchit, C., and T. Klaus. 2020. Determinants and impact of online reviews on product satisfaction. Journal of Internet Commerce 19 (1): 82–102.

Chen, Y.J., and A. Kirmani. 2015. Posting strategically: The consumer as an online media planner. Journal of Consumer Psychology 25 (4): 609–621.
Ramos, G.M., W. Daamen, and S. Hoogendoorn. 2014. A state-of-the-art review: Developments in utility theory, prospect theory and regret theory to investigate travellers’ behaviour in situations involving travel time uncertainty. *Transport Reviews* 34 (1): 46–67.

Rese, A., S. Schreiber, and D. Baier. 2014. Technology acceptance modeling of augmented reality at the point of sale: Can surveys be replaced by an analysis of online reviews? *Journal of Retailing and Consumer Services* 21 (5): 869–876.

Smith Brain Trust (SBT). 2022. The pandemic, the inflation and the way we shop now. https://www.rhsmith.umd.edu/research/pandemic-inflation-and-way-we-shop-now. Accessed 11 April 2022.

Stamolampros, P., and N. Korfiatis. 2018. Exploring the behavioral drivers of review valence. *International Journal of Contemporary Hospitality Management* 30 (10): 3083–3099.

Statista. 2021. Retail e-commerce sales worldwide. https://www.statista.com/statistics/379046/worldwide-retail-e-commerce-sales/. Accessed 5 April 2021.

Statista. 2022. E-commerce worldwide—Statistics & facts. https://www.statista.com/topics/871/online-shopping/. Accessed 10 April 2022.

Su, L., X. Yang, and Y. Huang. 2022. Tourists’ goal-directed behaviors: The influences of goal disclosure, goal commitment, and temporal distance. *Journal of Travel Research* 61 (4): 940–960.

Thaler, R.H. 2000. From Homo economicus to Homo sapiens. *The Journal of Economic Perspectives* 14: 133–141.

Thirumalai, S., and K.K. Sinha. 2005. Customer satisfaction with order fulfillment in retail supply chains: Implications of product type in electronic B2C transactions. *Journal of Operations Management* 23 (3–4): 291–303.

Trope, Y., and N. Liberman. 2010. Construal-level theory of psychological distance. *Psychological Review* 117 (2): 440.

Tversky, A., and D. Kahneman. 1992. Advances in prospect theory: Cumulative representation of uncertainty. *Journal of Risk and Uncertainty* 5 (4): 297–323.

UNCTAD. 2022. COVID-19 and e-commerce a global review. https://unctad.org/webflyer/covid-19-and-e-commerce-global-review. Accessed 10 April 2022.

Vakulenko, Y., P. Shams, D. Hellström, and K. Hjort. 2019. Online retail experience and customer satisfaction: The mediating role of last mile delivery. *The International Review of Retail, Distribution and Consumer Research* 29 (3): 306–320.

Walmsley, S., and A. Gilbey. 2020. Applying prospect theory to pilot weather-related decision-making: The impact of monetary and time considerations on risk taking behaviour. *Applied Cognitive Psychology* 34 (3): 685–698.

Wang, J.W., and C.H. Lin. 2021. Temporal distance and motivation matter: Effects of psychological distance and left-digit in price evaluation. *Current Psychology*. https://doi.org/10.1007/s12144-021-02359-2.

Williams, L.E., R. Stein, and L. Galguera. 2014. The distinct affective consequences of psychological distance and construal level. *Journal of Consumer Research* 40 (6): 1123–1138.

Wu, J., H. Zhao, and H. Chen. 2021. Coupons or free shipping? Effects of price promotion strategies on online review ratings. *Information Systems Research* 32 (2): 633–652.

Yang, Y., L. Wu, and W. Yang. 2018. Does time dull the pain? The impact of temporal contiguity on review extremity in the hotel context. *International Journal of Hospitality Management* 75: 119–130.

Yudkin, D.A., N. Liberman, C. Wakslak, and Y. Trope. 2020. Better off and far away: Reactions to others’ outcomes depends on their distance. *Organizational Behavior and Human Decision Processes* 156: 13–23.

Zeithaml, V.A., A. Parasuraman, and A. Malhotra. 2002. Service quality delivery through web sites: A critical review of extant knowledge. *Journal of the Academy of Marketing Science* 30 (4): 362.

Zhu, F., and X. Zhang. 2010. Impact of online consumer reviews on sales: The moderating role of product and consumer characteristics. *Journal of Marketing* 74 (2): 133–148.

**Publisher’s Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Dr. Prashanth Ravula is an Assistant Professor of Marketing at the NJCU School of Business, New Jersey City University. He received his Ph.D. (Marketing) from the University of Wisconsin - Milwaukee. Dr. Ravula's research interest is in Digital Marketing and has published his research in the *International Journal of Research in Marketing* and *Research in Transportation Economics*. 

Springer Nature