8-2023

Sharing Economy Last Mile Delivery: Three Essays Addressing Operational Challenges, Customer Expectations, and Supply Uncertainty

Nicolò Masorgo
University of Arkansas-Fayetteville

Sharing Economy Last Mile Delivery: Three Essays Addressing Operational Challenges, Customer Expectations, and Supply Uncertainty

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy in Business Administration

by

Nicolò Masorgo
University of Verona
Bachelor of Science in Economics, 2015
University of Verona
Master of Science in International Economics and Business Management, 2018

August 2023
University of Arkansas

This dissertation is approved for recommendation to the Graduate Council.

David D. Dobrzykowski, Ph.D.
Dissertation Co-Chair

Brian S. Fugate, Ph.D.
Dissertation Co-Chair

Rodney Thomas, Ph.D.
Committee Member

Iana Shaheen, Ph.D.
Committee Member

Christopher Tang, Ph.D.
Committee Member (External)

Brent Cunningham
Committee Member (External)

Abstract

Last mile delivery has become a critical competitive dimension facing retail supply chains. At the same time, the emergence of sharing economy platforms has introduced unique operational challenges and benefits that enable and inhibit retailers' last mile delivery goals. This dissertation investigates key challenges faced by crowdshipping platforms used in last mile delivery related to crowdsourced delivery drivers, driver-customer interaction, and customer expectations. We investigate the research questions of this dissertation through a multi-method design approach, complementing a rich archival dataset comprised of several million orders retrieved from a Fortune 100 retail crowdshipping platform, with scenario-based experiments. Specifically, the first study analyzes the impact of delivery task remuneration and operational characteristics that impact drivers' pre-task, task, and post-task behaviors. We found that monetary incentives are not the sole factor influencing drivers' behaviors. Drivers also consider the operational characteristics of the task when accepting, performing, and evaluating a delivery task. The second study examines a driver's learning experience relative to a delivery task and the context where it takes place. Results show the positive impact of driver familiarity on delivery time performance, and that learning enhances the positive effect. Finally, the third study focuses on how delivery performance shape customers' experience and future engagement with the retailer, examining important contingency factors in these relationships. Findings support the notion that consumers time-related expectations on the last mile delivery service influence their perceptions of the delivery performance, and their repurchase behaviors. Overall, this dissertation provides new insights in this emerging field that advance theory and practice.

Acknowledgements

I wish to thank my dissertation committee for guiding and supporting me throughout this journey: Dr. Dobrzykowski, Dr. Fugate, Dr. Shaheen, Dr. Thomas, Dr. Tang, and Mr. Cunningham. Each of you played a key role in shaping this dissertation and my present and future in Academia. David, you impacted my scholarship in several ways and always represented a great example of what success in Academia means. Without you accepting this challenge, this dissertation would not have been possible. Brian, besides offering your vast experience and unlimited expertise, you supported and directed my journey in several ways, believed in my capabilities, and offered several learning opportunities. Iana, working with a scholar with such positive energy and extended expertise is a pleasure. Since you joined the Department, your impact on us, Ph.D. students, has been tremendous. Rod, you impacted my journey since the first semester when I learned what research is in your seminar. Thank you for accepting this challenge and supporting me during the dissertation. Dr. Tang, it's a privilege and honor to have you serving on this committee. Your feedback and expertise have been fundamental to this dissertation's theoretical and methodological development. Last but surely not least, Brent, thank you for the incredible work and support. The rare opportunity to have real-world input influences a scholar's carrier.

I wish to thank those who impacted my journey and the past four years of my life. First, I want to thank my co-authors, who always supported my research and shared their knowledge and experience. I am also thankful to all the faculty, staff, and other Ph.D. students. In particular, Rob and Llord, it has been a pleasure to share these past four years together, learn from each other, and become great friends! Finally, I want to thank the Walton College of Business for influencing and nurturing my academic career. It's never easy to blend in with a different culture, unless you join the Walton College of Business.

I also wish to thank those who believed in me and saw my potential before joining the program. Ayman, Ilenia, Ivan, and Jon, thank you for impacting my life in several ways. Being exposed to such great scholars is a privilege of few. I am also proud to be the receiver of the award “In memoria del Prof. Antonio Borghesi” and want to thank the Department of Management at the University of Verona. Finally, I wish to thank my family, friends, and Ingrid. I could always count on their support. My family for believing in my dreams and life goals. I reserve a special thank you to the family members who have always guided and protected my journey. My Italian and International friends for constantly reminding me how much friendship matters in someone's life. Finally, Ingrid, for gifting my life with her love.

Table of Contents

I. Introduction 1

The crowdshipping context 6

Literature review overview and positioning the essays 8

Reference list 12

II. Essay 1: Driver behaviors in crowdshipping: An Income Opportunity Effect perspective 17

Introduction 17

Literature review: Sharing economy challenges in last mile delivery 21

Social Exchange Theory and Income Opportunity Effect 26

Hypotheses development 29

Empirical Setting and Data 35

Empirical Models 46

Results and Discussion 55

Conclusions, limitations, and directions for future research 73

Reference list 77

III. Essay 2: Leveraging driver's learning: The impact of driver's familiarity and order characteristics on delivery performance 90

Introduction 90

Literature review: Operational challenges in last mile delivery 94

Theoretical background 96

Hypotheses development 98

Empirical Setting and Data 100

Results 106

Discussion 109

Limitations and Future Research 112

Reference list 113

IV. Essay 3: Addressing consumers' preferences in last mile delivery: The impact of delivery time performance and order characteristics on customer outcomes 120

Introduction 120

Literature review: Consumers' challenges in last mile delivery 124

Expectancy Disconfirmation Theory 127

Hypotheses development 130

Methodology preview 136

Study 1 – Data and econometric models 137

Study 1 – Results 146

Study 1 – Discussion of results 156

Commitment Trust Theory 158

Study 2 – Experiment 160

Study 2 – Results 166

Study 2 – Discussion of results 177

Conclusions, limitations, and directions for future research 178

Reference list 182

V. Conclusion 200

Reference list 204

VI. Appendices 205

Appendix A – Essay 1 Robustness check and post-hoc analysis 205

Appendix B – Essay 2 Robustness check and post-hoc analysis 209

Appendix C – Essay 3: Study 1 sensitivity analysis 212

Appendix D – Essay 3: Study 2 experimental setup and vignette 218

Appendix F – Essay 3: Study 2 sensitivity analysis... 220

Reference list... 231

Appendix G – Institutional Review Board Protocol Approval 234

Table of Figures

Figure 1 – The order fulfillment process in crowdshipping delivery (adapted from Walmart.com) 6

Figure 2 – The order fulfillment process in crowdshipping delivery (adapted from Target.com) . 7

Figure 3 – Literature review and positioning of the essays 11

Figure 4 – Theoretical model essay 1 29

Figure 5 – Histogram for delivery tasks across calendar days........ 39

Figure 6 – Histogram for delivery tasks across days of the week 39

Figure 7 – Histogram for delivery tasks across clock hours........ 40

Figure 8 – Interaction between Remuneration and Density on Acceptance response time.......... 59

Figure 9 – Interaction between Remuneration and Expedited on Acceptance response time 59

Figure 10 – Interaction between Remuneration and Density on Service time 64

Figure 11 – Interaction between Remuneration and Expedited on Retention 68

Figure 12 – Kaplan-Meier survival function curve........ 70

Figure 13 – Cox proportional hazard regression model for remuneration per day........ 72

Figure 14 – Theoretical model essay 2 98

Figure 15 – Interaction between Driver-Customer Familiarity and Repeat Delivery Type (Table 15 Model 3)........ 108

Figure 16 – Interaction between Driver-Customer Familiarity and Urban (Table 15 Model 3) 109

Figure 18 – Distribution of Customer Rating 141

Figure 19 – Interaction effect of lateness and delivery window length on customer rating....... 150

Figure 20 – Interaction effect of earliness and delivery window length on customer rating...... 150

Figure 21 – Interaction effect of lateness and expedited delivery on customer rating 151

Figure 22 – Interaction effect of earliness and expedited delivery on customer rating............. 151

Figure 23 – Interaction effect of earliness and expedited delivery on repurchase behaviors 152

Figure 24 – Quadratic effect of earliness on repurchase behaviors with Johnson-Neyman lower 95% threshold (full red line), and upper 95% threshold (dashed red line) 153

Figure 25 – Kaplan-Meier curve for customers survival analysis 154

Figure 26 – Statistical diagram and notations for mediation analysis Process Model 4............. 171

Figure 27 – Statistical diagram and notations for mediation analysis Process Model 7............. 172

Figure 28 – Interaction effect of late and delivery window length on commitment................... 176

Figure 29 – Interaction effect of late and delivery window length on commitment................... 176

Figure 30 – Distribution of Review length across conditions (Review length: min = 1, max = 132) 231

Table of Tables

Table 1 – Descriptive Statistics 41

Table 2 – Correlations 41

Table 3 – Heckman's model first stage 49

Table 4 – Test for endogeneity 55

Table 5 – IV/2SLS results of the first stage 56

Table 6 – Test for endogeneity 57

Table 7 – IV/2SLS second stage results for Acceptance response time and delivery task remuneration 58

Table 8 – Complementary statistics to IV/2SLS second stage results for Acceptance response time and delivery task remuneration 60

Table 9 – IV/2SLS second stage results for Service time and delivery task remuneration 64

Table 10 – Complementary statistics to IV/2sls second stage results for Service time and delivery task remuneration 65

Table 11 – OLS estimations for Driver return time and daily remuneration 68

Table 12 – Cox proportional hazard regression model for remuneration per day 71

Table 13 – Descriptive Statistics 104

Table 14 – Correlations 104

Table 15 – Poisson Regression: The effect of Familiarity, Delivery type repetition, Repeat urban, and Interactions on Tardiness 108

Table 16 – Descriptive Statistics 140

Table 17 – Correlations 140

Table 18 – IV/2SLS results of the first stage 147

Table 19 – Test for endogeneity 148

Table 20 – IV/2SLS second stage. Direct effects of lateness and earliness on customer outcomes 148

Table 21 – IV/2SLS second stage. Moderation effects of delivery window length and expedited delivery 149

Table 22 – Cox proportional hazard regression model for lateness and earliness 155

Table 23 – Summary of participants demographisc and experimental procedures 163

Table 24 – Experimental checks 164

Table 25 – Confirmatory factor analysis and scale items statistics 167

Table 26 – Pairwise correlations of substantive variables 168

Table 27 – Operationalization of covariates included in the mediation analysis 168

Table 28 – Common method bias tests 169

Table 29 – Mediation results of Process Model 4 with parallel mediation for the effect of Late (L) and Early (E) on Satisfaction with the retailer (SAT) and Repurchase intentions (REP) via Trust (T) and Commitment (C) 171

Table 30 – Moderated Mediation results for the effect of On-time vs Late on Satisfaction with the retailer (SAT) and Repurchase Intentions (REP), via Trust (T) and Commitment (C), moderated by Window Length (WL) 173

Table 31 – Moderated Mediation results for the effect of On-time vs Early on Satisfaction with the retailer (SAT) and Repurchase Intentions (REP), via Trust (T) and Commitment (C), moderated by Window Length (WL) 174

Table 32 – Moderated Mediation results for the effect of Early vs Late on Satisfaction with the retailer (SAT) and Repurchase Intentions (REP), via Trust (T) and Commitment (C), moderated by Window Length (WL) 175

Table 33 – Robustness check results of Table 7 with imr computed with zip code fixed effects 205

Table 34 – Robustness check results of Table 7 with Poisson regression 206

Table 35 – Robustness check results of Table 7 with imr computed with zip code fixed effects 207

Table 36 – Cox proportional hazard regression model specifying a Weibull distribution for remuneration per day 208

Table 37 – Aalen’s additive hazard model estimations for remuneration per day 208

Table 38 – Robustness check results of Table 4 – Estimations with Logit Regression and Binary Dependent Variable Delivery Performance (1 = late; 0 otherwise) 209

Table 39 – Robustness check results of Table 4 – Poisson Regression Estimations including Restart 210

Table 40 – Robustness check results of Table 4 – Poisson Regression Estimations with moderators Binary Variables Unattended and Urban 211

Table 41 – Results estimation models to address endogeneity for sample bias 212

Table 42 – Estimation results for Customer Satisfaction including Inverse Mills Ratio as control variable 214

Table 43 – Estimation results for Repurchase Intentions including Inverse Mills Ratio as control variable 214

Table 44 – Estimation results for Customer Satisfaction using Ordered Logit Regression 215

Table 45 – Estimation results for Repurchase Intentions including Customer Satisfaction as control variable 216

Table 46 – Estimation results for Repurchase Intentions including zip code fixed effects 217

Table 47 – Experimental cues module delivery performance - Message in the email screenshot 219

Table 48 – Mediation results of Process Model 4 without control variables 220

Table 49 – Moderated Mediation results of On-time vs Late without control variables 221

Table 50 – Moderated Mediation results of On-time vs Early without control variables 222

Table 51 – Moderated Mediation results of On-time vs Early without control variables 223

Table 52 – Mediation results of Process Model 4 with full sample (N = 549) 225

Table 53 – Moderated Mediation results of On-time vs Late with full sample (N = 549) 226

Table 54 – Moderated Mediation results of On-time vs Early with full sample (N = 549) 227

Table 55 – Moderated Mediation results of On-time vs Early with full sample (N = 549) 228

Table 56 – Moderation results of Process Model 5 with moderation in the direct effect of Lateness (L) and Earliness (E) on Repurchase intentions (REP) 229

Table 57 – Results for Post-hoc analysis 2: The effect of Late and Early on review mood and review length.. 230

I. Introduction

Last mile delivery represents the last portion of a retail supply chain that transfers packages from a distribution center to a receiver (Esper et al., 2003; Deng, Fang, & Lim, 2021). Despite being one of the most important success factors of the order fulfillment process that shape customer's outcomes, last mile delivery still presents a plethora of unresolved challenges that result in an inefficient and ineffective service (Lee & Whang, 2001; Boyer, Prud'homme, & Chung, 2009; Ishfaq et al., 2016; Nguyen et al., 2019; Ross, 2021; Wang, Rabinovich, & Guda, 2022). Last mile delivery is one of the most expensive aspects of delivery logistics (Fatehi & Wagner, 2022).

Last mile delivery can be inefficient primarily because it involves high volumes of individual small purchase quantities, often with single packing, and unit delivery to end-consumer (Lim & Winkenbach, 2019; Dolan, 2022). It follows that last mile delivery is often performed using a dedicated fleet that delivers low-value items on less-than-truckload (Boyer & Hult, 2005; Ishfaq et al., 2016; Lim & Winkenbach, 2019). These complexities result in an increasing cost associated to the last mile, which accounts for 53% of the total shipping costs (Punakivi, Yrjölä, & Holmström, 2001; Boyer et al., 2009; Jacobs et al., 2019; Dolan, 2022). In many instances, last mile delivery is also ineffective. On-time deliveries represent a key challenge for retail supply chains, which struggle to improve supply chain flows and coordinate the key actors performing the delivery (Awaysheh et al., 2021; Liu, He, & Shen, 2021; Mao et al., 2022).

New sharing economy business models have addressed such challenges by outsourcing deliveries tasks to the crowd, and managing deliveries through online platform. One of the most important business models is crowdshipping. Crowdshippng is the practice in which organizations or persons who needs to transport an item are connecting through a sharing economy service platform with individuals of the crowd that are willing to perform the delivery (Dayarian &

Savelsbergh, 2020). Thanks to its flexibility and the proximity of fleet capacity, retail supply chains have been more and more relying on crowdshipping, which has proved useful to handle single order expedited deliveries (e.g., food or groceries same-day delivery) (Dayarian & Savelsbergh, 2020; Dayarian & Pazour, 2022). Notable examples include Amazon launching the Amazon Flex program, Walmart using crowdsourced drivers through Spark and GoLocal, and many local retailers adopting the services offered by crowdshipping platforms – sharing economy service companies utilizing cloud-based technologies to match demand and supply (Apte & Davis, 2019; CSCMP, 2022; Fatehi & Wagner, 2022).

Recent trends have exacerbated these criticalities. First, the expansion of e-commerce has increased the volume of packages to deliver, creating pressure on the system, especially after the outbreak of the Covid-19 pandemic (Ketchen & Craighead, 2020; CSCMP, 2022; Delasay, Jain, & Kumar, 2022; Lyu & Teo, 2022). The first challenge relates to the scarcity of delivery drivers performing deliveries, which further increases the tension in last mile delivery to fulfill orders while trying to manage increasing volumes and customers' expectations (Straight, 2021). The second trends relates to driver's learning. While last mile delivery in a crowdshipping context is considered a trivial activity, performing high quality deliveries that include on-time performance and customer satisfaction can present challenges for unprofessional drivers. Finally, the third trend relates to increasing consumers' expectations of the delivery service (Hübner, Kuhn, & Wollenburg, 2016; Daugherty, Bolumole, & Grawe, 2019). It is essential for a retail supply chain to provide customers with on-time, fast, precise, and free-of-charge deliveries with minimal fulfillment time that meets their convenience, pushing the delivery service time performance to achieve both speed and punctuality (Fisher, Gallino, & Xu, 2015; Gawor & Hoberg, 2019; Mangiaracina et al., 2019). Retailers offer several delivery speed options to win online customer

orders (e.g., same-day delivery) (Ishfaq et al., 2016; Peinkofer, Schwieterman, & Miller, 2020), increasing the pressure on an already-thin profit margin fulfillment service that is yet indispensable, especially for large retailers (Dayarian & Savelsbergh, 2020; Kammerer et al., 2020; Stroh, Erera, & Toriello, 2022).

Several research streams have been investigating these challenges. The literature has identified key sources of capacity uncertainty stemming from sharing economy business models in operations management and last mile delivery (Benjaafar & Hu, 2020; Dong & Ibrahim, 2020). Often, crowdshipping platforms stimulate supply through dynamic surge pricing– the practice of increasing prices during times of high demand (Cachon, Daniels, & Lobel, 2017), especially when demand volume exceeds the supply (Castillo et al., 2022a). However, it has not looked at operational characteristics influencing driver's behaviors.

In addition, crowdshipping drivers often drop this gig job after few months of practicing it (Bernstein, DeCroix, & Keskin, 2021). Thus, their learning experience and expertise gains should not be tangible in this space. Crowdshipping platforms struggle to recruit and nurture drivers (Apte & Davis, 2019). However, typically delivery drivers assigned to the same route, because familiarity in the delivery route improve their knowledge of the delivery task, satisfies customers' needs, and increase drivers' comfort and retention (Keller, 2002; Ulmer et al., 2021). Thus, through driver's learning and familiarity, crowdshipping platforms help improving such outcomes.

Finally, the third body of literature refers to the literature on logistics service quality and physical distribution service quality, which traditionally constitute a core part of last mile delivery literature (Nguyen et al., 2019). This body of research identified three key dimensions of service quality that impact customers' outcomes, namely operational, economic, and relational (Mentzer, Flint, & Hult, 2001; Stank et al., 2003). Operational aspects of service quality are crucial for

successful delivery and include on-time and fast deliveries, number of shipping options, product availability, ability to track orders, and product handling (Rabinovich & Evers, 2003; Rabinovich, 2004; Rabinovich & Bailey, 2004; Rabinovich, Rungtusanatham, & Laseter, 2008; Rao, Griffis, & Goldsby, 2011; Rao, Rabinovich, & Raju, 2014). The economic dimension refers to the shipping fees, an aspect that shapes customer's expectations and impacts customers' satisfaction relative to the level of service quality (Lewis, Singh, & Fay, 2006; Gümüş et al., 2013; Ma, 2017; Nguyen et al., 2019; Barker & Brau, 2020; Tokar, Williams, & Fugate, 2020). Finally, the relational dimension refers to the interaction occurring in the order fulfillment process between the focal firm and the customer (Davis-Sramek, Mentzer, & Stank, 2008). While it is deemed important, this dimension has received little attention in the Business-to-Customer, which research focuses primarily on the key role of the delivery driver (Ta, Esper, & Rossiter, 2018; Daugherty et al., 2019; Peinkofer et al., 2020; Puram et al., 2021; Castillo et al., 2022b)

This dissertation investigates the challenges that a crowdsourcing platform faces when providing the last mile delivery service. This dissertation is comprised of three studies. The first essay examines how delivery task remuneration and operational characteristics influence driver's behaviors pre-, during-, and post-task. Specifically, increasing the deliver task remuneration reduces supply-side uncertainty, because drivers will be more likely to accept delivery offers for high volume batches, thus reducing supply uncertainty. However, drivers' preferences related to a delivery offer may also depend on the characteristics of the delivery. Hence this essay investigates how delivery remuneration influences service provider behaviors, and how delivery characteristics moderates this relationship. Econometric analysis reveal that monetary incentives are not the sole factor influencing drivers' behaviors. Tasks with greater delivery density can reduce the acceptance response time up to 5 minutes, and the service time by an additional 7 minutes. This

results, combined with the fact that monetary incentives play the major role in driver retention, suggest that supply uncertainty reduces as platforms concentrate their effort in recruiting and retaining drivers that show a longer commitment to the platform.

The second essay focuses on driver's familiarity and how learning opportunities improve delivery outcomes, as well as the contingency factors that influence this relationship. Findings from econometric analysis show a significant improvement in delivery time performance when drivers gain familiarity with delivering to a customer and after repeating the same delivery type. Finally, while urban deliveries present worse delivery time performance, the interplay with familiarity does not seem to significantly impact the performance. That is, the delivery context does not affect how familiarity improves delivery time performance.

Finally, the third essay study investigates how delivery performance and contingency factors affecting customers service expectations, such as delivery window length and expedited delivery, can impact customer's outcomes. Results from econometric analyses combined with a scenario-based experiment show the importance of meeting the delivery window: both earliness and lateness present a negative effect on customer outcomes. This study also shows the trade-off generated by delivery window length. A shorter delivery window increases customer expectations that are more difficult to meet, but improves the service experience. A longer window decreases expectations, but dissatisfy customers. In addition, to reconcile the result that early deliveries would improve loyalty but decrease repurchase behaviors, an experiment clarifies the effect of delivery failures affect the customer-retailer relationship. Overall, this study informs theory and retailers on the importance of managing delivery window slots and delivery performance, as an effort to effectively handle consumers time-related expectations on the last mile delivery service.

This research endeavor offers several theoretical contributions relative to the crowdshipping and last mile delivery literature, and to the theoretical background used in these studies. Further, this dissertation offers several managerial contributions and offers guidelines on managing the last mile in crowdshipping.

The crowdshipping context

To address the literature gaps, we partnered with a crowdshipping platform – that we refer to using the pseudonym *Alpha* – delivering for a Fortune 100 retailer. In the evolving US last-mile delivery market, *Alpha* is one of the most promising emerging white-label players offering next-day and two-day delivery service from local fulfillment hubs and stores (CSCMP, 2022).

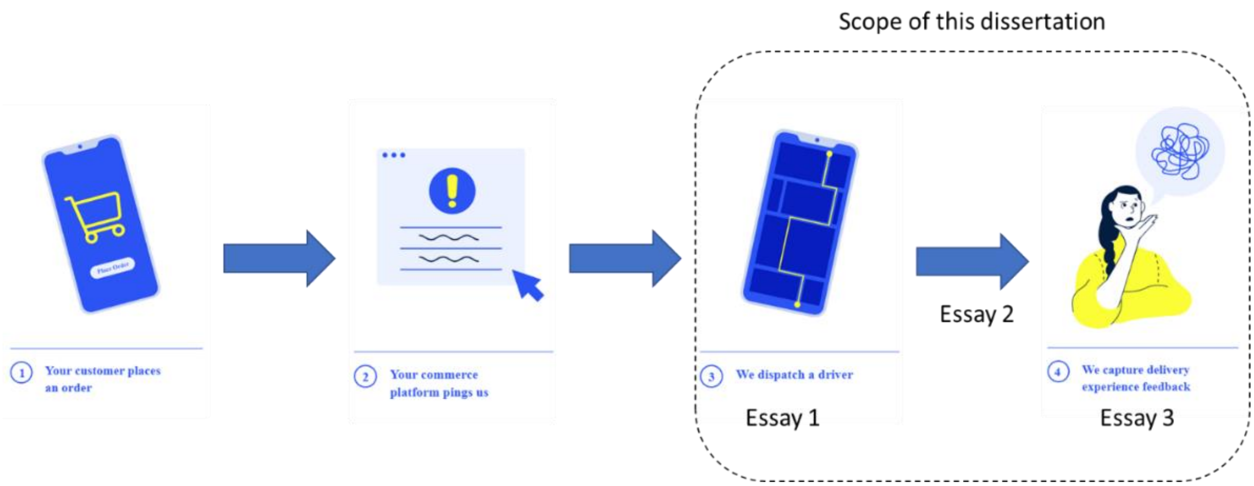


Figure 1 – The order fulfillment process in crowdshipping delivery (adapted from Walmart.com)

The supply chain in which *Alpha* operates comprises of four major players: the online crowdshipping platform (*Alpha*), the retailers, and the crowd, which includes customers and delivery drivers. While the order fulfillment process in crowdshipping follows a traditional linear framework of forward logistics (Rabinovich & Bailey, 2004), the material, information, and financial flows present some unique characteristics. Figures 1 and 2 present exemplars of the order

fulfillment process in crowdshipping delivery for two prominent retailers. This framework resembles those found in the crowdshipping literature (Mao et al., 2022).

Figure 2 – The order fulfillment process in crowdshipping delivery (adapted from Target.com)

In the order fulfillment process, defined as a set of interrelated activities from the point of a customer's purchase decision to the point a product is delivered to the customer (Croxton, 2003), as a white label service provider, *Alpha* manages only the final part of the fulfillment, specifically the activities of last-mile delivery. Thus, in the first step of the fulfillment process, the customer places an order to the retailer, who then communicate with *Alpha* relative to the characteristics of the order through an integrated information system (IP integration). *Alpha* dispatches a driver by broadcasting the delivery task with the remuneration offer to the drivers in the designated delivery zone where the pick-up point belongs. Finally, the *Alpha* driver completes the delivery task, while *Alpha* collects customers' feedback on the delivery performance. A customer places an order to the retailer, paying and informing about the desired delivery. The retailer shared with *Alpha*, who dispatches the driver by communicating delivery details and, once the delivery completed, the remuneration. *Alpha* optimized the delivery task composition in terms of routing, delivery service,

and timing. Hence, the number and which orders are included in the delivery task follow a specific optimization algorithm. The driver commutes to the retailer's facility, collects the orders, and delivers to the final customers' destination. The customer voluntarily tips the driver and provides her/his feedback.

We completed the collection of a large dataset from *Alpha*. The dataset includes delivery operational data, including time stamps, orders characteristics, and geolocation data on pickup and dropoff points, transactional data relative to customers, including customers' outcomes metrics, demographic data on the drivers, and dispatching store characteristics. Hence, referring to Figure 1, the scope of this dissertation is from step 3 (we dispatch the driver) to 4 (we capture delivery experience feedback). These steps represent last-mile delivery operations.

Literature review overview and positioning the essays

The three studies of this dissertation are grounded on an extensive literature review on sharing economy and last mile delivery. To improve the flow of the three essays, we report each body of literature within each essay. In this section, we report methodological details on the literature review and an overview of the results.

Literature review material collection

We retrieved the material for the literature review by searching for relevant keywords in the main journal of interest. Specifically, given the focus of this work is sharing economy and last mile delivery, we searched for the keywords "sharing economy" "crowdshipping" "ridesharing" "gig economy" "last mile delivery". We selected a detailed journal list to ensure comprehensiveness in both the theoretical, notional, and methodological background. First, we focused on scientific journals discussing logistics, supply chain management, and operations management issues. We included Journal of Supply Chain Management, Journal of Business Logistics, International

Journal of Logistics Management, Supply Chain Management: An International Journal, International Journal of Physical Distribution and Logistics Management, Management Science, Manufacturing and Service Operations, Decision Sciences, Transportation Journal. However, upon reviewing the journal articles, we realized the need to extend the research to management journals for a in-depth theoretical background on sharing economy, and to information systems journals to gain a clearer understanding of the technology behind sharing economy. Thus, we expanded the research to Academy of Management Review, Academy of Management Journal, Academy of Management Annals, Academy of Management Discoveries, Strategic Management Journal, Administrative Science Quarterly, Journal of Management, California Management Review, MIS Quarterly, and MIS Quarterly Executive. We retrieved a total of 179 journal articles.

Literature overview and positioning of the three essays

The literature review reveals several body of research in the space of last mile delivery and sharing economy. In this section, we direct the reader relative to the positioning of each essay within a specific body of literature (see Figure 3). Each essay will contain a relevant literature review based on the respective research questions. Crowshipping presents characteristics of capacity and flexibility to address such last mile delivery challenges. However, crowdshipping suffers from delivery drivers' supply-side uncertainty, which does not allow retailer supply chains to keep up with increasing volumes as it reduces capacity, meet customers' expectations, and offer an economically feasible solution. Attracting and retaining drivers seems a key element for the consistent usage of crowdshipping in last mile delivery. A way to attract drivers is to provide them with a regular stream of income (Cameron, 2022). This can be achieved through increasing delivery volumes, because drivers can find the task engaging and remunerative for an unqualified job. However, operational characteristics of the delivery task may affect driver's behaviors.

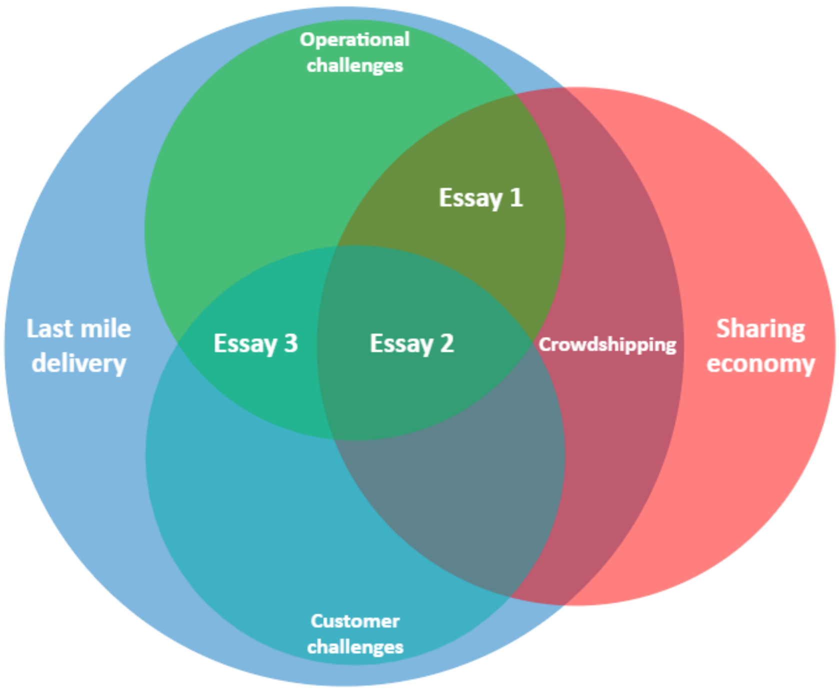
Drivers may not think that the delivery is profitable because as volume increases, number of drop-off points increase, unless the delivery is focused on a limited geographical area. Drop-off density represents one of the key elements to decrease drivers' supply-side uncertainty (Qi et al., 2018). Other delivery characteristics may also influence driver's supply-side uncertainty. Hence, the first essay is located at the intersection between last mile delivery operational challenges and crowdsourced delivery.

Second, the literature reveals several operational challenges in last mile delivery, which can be addressed by crowdshipping. On the one hand, the operational challenges related to meeting the increasing delivery demand, expanding the offer of delivery options, and achieving profitability, have been addressed through optimization models and last mile delivery innovations integration. An alternative solution could be to focus on managing volumes and drop-off density to implement economies of scale in last mile delivery, as prescribed by (Qi et al., 2018). Finally, research suggests that in crowdshipping, driver's familiarity with the route can increase the delivery performance and the customer outcome. Hence, the second essay is located at the intersection between crowdsourced delivery and customers' challenges.

Finally, consumers' expectations of a greater level of service requires additional capacity that often is not available to provide an on-demand, cheap, and precise service. The literature studied how the perceived service quality helps to understand customers. However, the literature overlooks at understanding customers' outcomes relative to such high expectations on last mile delivery. Indeed, A relative unexplored area relates to how consumers have changed their delivery preferences and how these preferences impact customers outcomes. Customers want faster, on-time, precise, and punctual deliveries. For example, is the customer more tolerant when an on-demand 2-hour delivery is requested? This is particularly important when trying to provide a

punctual and precise delivery, meeting the delivering window, while addressing the customers' requirements. Hence, the third essay is located between last mile delivery operational challenges and customers' outcomes.

Figure 3 – Literature review and positioning of the essays



Reference list

Apte, U. M., & Davis, M. M. 2019. Sharing economy services: Business model generation. *California Management Review*, 61(2): 104–131.

Awaysheh, A., Frohlich, M. T., Flynn, B. B., & Flynn, P. J. 2021. To err is human: Exploratory multilevel analysis of supply chain delivery delays. *Journal of Operations Management*, 67(7): 882–916.

Barker, J. M., & Brau, R. I. 2020. Shipping surcharges and LSQ: pricing the last mile. *International Journal of Physical Distribution and Logistics Management*, 50(6): 667–691.

Benjaafar, S., & Hu, M. 2020. Operations management in the age of the sharing economy: What is old and what is new? *Manufacturing and Service Operations Management*, 22(1): 93–101.

Bernstein, F., DeCroix, G. A., & Keskin, N. B. 2021. Competition between two-sided platforms under demand and supply congestion effects. *Manufacturing and Service Operations Management*, 23(5): 1043–1061.

Boyer, K. K., & Hult, G. T. M. 2005. Extending the supply chain: Integrating operations and marketing in the online grocery industry. *Journal of Operations Management*, 23(6): 642–661.

Boyer, K. K., Prud'homme, A. M., & Chung, W. 2009. The Last Mile Challenge: Evaluating the Effects Of Customer Density and Delivery Window Patterns. *Journal of Business Logistics*, 30(1): 185–201.

Cachon, G. P., Daniels, K. M., & Lobel, R. 2017. The Role of Surge Pricing on a Service Platform with Self-Scheduling Capacity. *Manufacturing & Service Operations Management*, 19(3): 368–384.

Cameron, L. D. 2022. "Making Out" While Driving: Relational and Efficiency Games in the Gig Economy. *Organization Science*, 33(1): 231–252.

Castillo, V. E., Bell, J. E., Mollenkopf, D. A., & Stank, T. P. 2022a. Hybrid last mile delivery fleets with crowdsourcing: A systems view of managing the cost-service trade-off. *Journal of Business Logistics*, 43(1): 36–61.

Castillo, V. E., Mollenkopf, D. A., Bell, J. E., & Esper, T. L. 2022b. Designing technology for on-demand delivery: The effect of customer tipping on crowdsourced driver behavior and last mile performance. *Journal of Operations Management*, 68(5): 424–453.

Croxton, K. L. 2003. The Order Fulfillment Process. *The International Journal of Logistics Management*, 14(1): 19–32.

CSCMP. 2022. *33rd Annual State of Logistics Report*. *Council of supply chain management professionals*.

Daugherty, P. J., Bolumole, Y., & Grawe, S. J. 2019. The new age of customer impatience: An agenda for reawakening logistics customer service research. *International Journal of Physical Distribution & Logistics Management*, 49(1): 4–32.

Davis-Sramek, B., Mentzer, J. T., & Stank, T. P. 2008. Creating consumer durable retailer customer loyalty through order fulfillment service operations. *Journal of Operations Management*, 26(6): 781–797.

Dayarian, I., & Pazour, J. 2022. Crowdsourced order-fulfillment policies using in-store customers. *Production and Operations Management*, 1–32.

Dayarian, I., & Savelsbergh, M. 2020. Crowdshipping and Same-day Delivery: Employing In-store Customers to Deliver Online Orders. *Production and Operations Management*, 29(9): 2153–2174.

Delasay, M., Jain, A., & Kumar, S. 2022. Impacts of the COVID-19 pandemic on grocery retail operations: An analytical model. *Production and Operations Management*, (September 2020).

Deng, Q., Fang, X., & Lim, Y. F. 2021. Urban Consolidation Center or Peer-to-Peer Platform? The Solution to Urban Last-Mile Delivery. *Production and Operations Management*, 30(4): 997–1013.

Dolan, S. 2022. The challenges of last mile delivery logistics and the tech solutions cutting costs in the final mile. *Business Insider*. Retrieved November 1, 2022, from https://www.businessinsider.com/last-mile-delivery-shipping-explained

Dong, J., & Ibrahim, R. 2020. Managing supply in the on-demand economy: Flexible workers, full-time employees, or both? *Operations Research*, 68(4): 1238–1264.

Esper, T. L., Jensen, T. D., Turnipseed, F. L., & Burton, S. 2003. The Last Mile: an Examination of Effects of Online Retail Delivery Strategies on Consumers. *Journal of Business Logistics*, 24(2): 177–203.

Fatehi, S., & Wagner, M. R. 2022. Crowdsourcing Last-Mile Deliveries. *Manufacturing & Service Operations Management*, 24(2): 791–809.

Fisher, M., Gallino, S., & Xu, J. J. 2015. The Value of Rapid Delivery in Online Retailing. *SSRN Electronic Journal*, 56(5): 732–748.

Gawor, T., & Hoberg, K. 2019. Customers' valuation of time and convenience in e-fulfillment. *International Journal of Physical Distribution and Logistics Management*, 49(1): 75–98.

Gümüş, M., Li, S., Oh, W., & Ray, S. 2013. Shipping fees or shipping free? a tale of two price partitioning strategies in online retailing. *Production and Operations Management*, 22(4): 758–776.

Hübner, A., Kuhn, H., & Wollenburg, J. 2016. Last mile fulfilment and distribution in omni-channel grocery retailing: A strategic planning framework. *International Journal of Retail and Distribution Management*, 44(3): 228–247.

Ishfaq, R., Defee, C. C., Gibson, B. J., & Raja, U. 2016. Realignment of the physical distribution process in omni-channel fulfillment. *International Journal of Physical Distribution and Logistics Management*, 46(6–7): 543–561.

Jacobs, K., Warner, S., Rietra, M., Mazza, L., Buvat, J., Khadikar, A., Cherian, S., & Khemka, Y.

2019. *The Last-Mile Delivery Challenge*. Retrieved from https://www.capgemini.com/research/the-last-mile-delivery-challenge/?utm_source=linkedin&utm_medium=social&utm_campaign=socialIN&utm_content=CRI_Report_Last-Mile Delivery_link

Kammerer, W., Chinnareddy, S., Mahoney, K., & Neff, D. 2020. *The last mile: Enabling last mile through speed and flexibility upstream*.

Keller, S. B. 2002. Driver Relationships With Customers and Driver Turnover: Key Mediating Variables Affecting Driver Performance in the Field. *Journal of Business Logistics*, 23(1): 39–64.

Ketchen, D., & Craighead, C. W. 2020. Research at the Intersection of Entrepreneurship, Supply Chain Management, and Strategic Management: Opportunities Highlighted by COVID-19. *Journal of Management*, 46(8): 1330–1341.

Lee, H. L., & Whang, S. 2001. Winning the Last Mile of E-Commerce. *MIT Sloan Management Review*, 42(4): 54–62.

Lewis, M., Singh, V., & Fay, S. 2006. An empirical study of the impact of nonlinear shipping and handling fees on purchase incidence and expenditure decisions. *Marketing Science*, 25(1): 51–64.

Lim, S. F. W. T., & Winkenbach, M. 2019. Configuring the last-mile in businessto- consumer e-retailing. *California Management Review*, 61(2): 132–154.

Liu, S., He, L., & Shen, Z. J. M. 2021. On-time last-mile delivery: Order assignment with travel-time predictors. *Management Science*, 67(7): 4095–4119.

Lyu, G., & Teo, C.-P. 2022. Last Mile Innovation: The Case of the Locker Alliance Network. *Manufacturing & Service Operations Management*, (Forthcoming).

Ma, S. 2017. Fast or free shipping options in online & Omni-channel retail? The mediating role of uncertainty on satisfaction & purchase intentions. *International Journal of Logistics Management*, 28(4): 1099–1122.

Mangiaracina, R., Perego, A., Seghezzi, A., & Tumino, A. 2019. Innovative solutions to increase last-mile delivery efficiency in B2C e-commerce: a literature review. *International Journal of Physical Distribution and Logistics Management*, 49(9): 901–920.

Mao, W., Ming, L., Rong, Y., Tang, C. S., & Zheng, H. 2022. On-Demand Meal Delivery Platforms: Operational Level Data and Research Opportunities. *Manufacturing & Service Operations Management*, (May).

Mentzer, J. T., Flint, D. J., & Hult, G. T. M. 2001. Logistics service quality as a segment-customized process. *Journal of Marketing*, 65(4): 82–104.

Nguyen, D. H., de Leeuw, S., Dullaert, W., & Foubert, B. P. J. 2019. What Is the Right Delivery Option for You? Consumer Preferences for Delivery Attributes in Online Retailing. *Journal of Business Logistics*, 40(4): 299–321.

Peinkofer, S. T., Schwieterman, M. A., & Miller, J. W. 2020. Last-mile delivery in the motor-

carrier industry: A panel data investigation using discrete time event history analysis. *Transportation Journal*, 59(3): 129–164.

Punakivi, M., Yrjölä, H., & Holmström, J. 2001. Solving the last mile issue: Reception box or delivery box? *International Journal of Physical Distribution and Logistics Management*, 31(6): 427–439.

Puram, P., Gurumurthy, A., Narmetta, M., & Mor, R. S. 2021. Last-mile challenges in on-demand food delivery during COVID-19: understanding the riders' perspective using a grounded theory approach. *The International Journal of Logistics Management*.

Qi, W., Li, L., Liu, S., & Shen, Z. M. 2018. Shared Mobility for Last-Mile Delivery: Design, Operational Prescriptions, and Environmental Impact. *Manufacturing & Service Operations Management*, 20(4): 737–751.

Rabinovich, E. 2004. Internet Retailing Intermediation: a Multilevel Analysis of Inventory Liquidity and Fulfillment Guarantees. *Journal of Business Logistics*, 25(2): 139–169.

Rabinovich, E., & Bailey, J. P. 2004. Physical distribution service quality in Internet retailing: service pricing, transaction attributes, and firm attributes. *Journal of Operations Management*, 21(6): 651–672.

Rabinovich, E., & Evers, P. T. 2003. Product Fulfillment in Supply Chains Supporting Internet-Retailing Operations. *Journal of Business Logistics*, 24(2): 205–236.

Rabinovich, E., Rungtusanatham, M., & Laseter, T. M. 2008. Physical distribution service performance and Internet retailer margins: The drop-shipping context. *Journal of Operations Management*, 26(6): 767–780.

Rao, S., Griffis, S. E., & Goldsby, T. J. 2011. Failure to deliver? Linking online order fulfillment glitches with future purchase behavior. *Journal of Operations Management*, 29(7–8): 692–703.

Rao, S., Rabinovich, E., & Raju, D. 2014. The role of physical distribution services as determinants of product returns in Internet retailing. *Journal of Operations Management*, 32(6): 295–312.

Ross, S. 2021. Seven Last-Mile Delivery Challenges, and How to Solve Them. *Supply Chain Brain*. Retrieved from https://www.supplychainbrain.com/blogs/1-think-tank/post/32800-last-mile-delivery-challenges-and-how-to-solve-them

Stank, T. P., Goldsby, T. J., Vickery, S. K., & Katrina, S. 2003. Logistics Service Performance: Estimating Its Influence On Market Share. *Journal of Business Logistics*, 24(1): 27–55.

Straight, B. 2021. Last-mile delivery firms call lack of drivers their main concern. *FreightWaves*. Retrieved from https://www.freightwaves.com/news/last-mile-delivery-firms-call-lack-of-drivers-their-main-concern

Stroh, A. M., Erera, A. L., & Toriello, A. 2022. Tactical Design of Same-Day Delivery Systems. *Management Science*, 68(5): 3444–3463.

Ta, H., Esper, T. L., & Rossiter, A. 2018. Designing crowdsourced delivery systems : The effect of driver disclosure and ethnic similarity. *Journal of Operations Management*, 60(1): 19–33.

Tokar, T., Williams, B. D., & Fugate, B. S. 2020. I Heart Logistics—Just Don't Ask Me to Pay For It: Online Shopper Behavior in Response to a Delivery Carrier Upgrade and Subsequent Shipping Charge Increase. *Journal of Business Logistics*, 41(3): 182–205.

Ulmer, M. W., Thomas, B. W., Campbell, A. M., & Woyak, N. 2021. The restaurant meal delivery problem: Dynamic pickup and delivery with deadlines and random ready times. *Transportation Science*, 55(1): 75–100.

Wang, L., Rabinovich, E., & Guda, H. 2022. An analysis of operating efficiency and policy implications in last-mile transportation following Amazon's integration. *Journal of Operations Management*, 1–27.

II. Essay 1: Driver behaviors in crowdshipping: An Income Opportunity Effect perspective

Introduction

A global research study reveals that almost 90% of retailers expect to use crowdshipping to deliver single orders by 2028 (Zebra Technologies, 2018). Crowdshippng is the practice in which organizations or persons who needs to transport an item are connecting through a sharing economy service platform with individuals of the crowd that are willing to perform the delivery (Dayarian & Savelsbergh, 2020). Thanks to its flexibility and the proximity of fleet capacity, retail supply chains have been more and more relying on crowdshipping, which has proved useful to handle single order expedited deliveries (e.g., food or groceries same-day delivery) (Dayarian & Savelsbergh, 2020; Dayarian & Pazour, 2022). Notable examples include Amazon launching the Amazon Flex program, Walmart using crowdsourced drivers through Spark and GoLocal, and many local retailers adopting the services offered by crowdshipping platforms – sharing economy service companies utilizing cloud-based technologies to match demand and supply (Apte & Davis, 2019; CSCMP, 2022; Fatehi & Wagner, 2022).

Despite its success, crowdshipping presents several challenges that increase uncertainty in last mile delivery operations (Benjaafar & Hu, 2020). A major challenge relates to understanding service providers behaviors pre-task, task, and post-task delivery (Cachon, Daniels, & Lobel, 2017; Ta, Esper, & Tokar, 2021; Dayarian & Pazour, 2022). In crowdshipping, online platforms seek to match an on-demand delivery task with service providers (Fatehi & Wagner, 2022). A delivery task corresponds to the offer that a platform broadcasts to the poo of crowdsourced drivers in an assigned delivery zone. In crowdshipping, the delivery task comprises a batch of single or more orders that the crowdsourced drivers, upon acceptance, will deliver on a specific route (Arslan et

al., 2019; Dayarian & Savelsbergh, 2020; Castillo et al., 2022b). Crowdsourced delivery service providers (hereafter referred to as "drivers") are independent contractors who deliver utilizing their own vehicle and independently determine their working schedule (Carbone, Rouquet, & Roussat, 2017). Pre-task behaviors refer to the driver accepting a delivery task. In crowdshipping, service supply is heterogenous in space, time, and capacity, as drivers decide where and when to work, as well as the type and quality of work (Benjaafar & Hu, 2020). This implies that crowdshipping platforms often deal with supply-side uncertainty, because it is not possible to exactly predict the number of drivers that will be available in a specific location to deliver an on-demand order (Dong & Ibrahim, 2020). During-task behaviors refers to the service commitment of the driver toward the performance of the delivery task (Dayarian & Pazour, 2022). Finally, post-task behaviors refer to driver feedback and retention (Ta et al., 2021).

Prior literature investigated driver behaviors by primarely focusing on supply-side uncertainty in the crowdshipping. A stream of literature studied how retail supply chains create hybrid systems to compensate the uncertainty of crowdshipping service supply with a private fleet (Castillo et al., 2018). This literature focused on compensation schemes for private fleet vs crowdsourced drivers to optimally design the delivery system (Dai & Liu, 2020; Castillo et al., 2022a). However, such hybrid fleet composition is optimal only under the circumstances that the size and mix between delivery fleets does not affect the cost-service trade-off (Castillo et al., 2022a). In other words, the hybrid fleet offers an advantage when the retailer is capable of matching fast shipping orders with crowdshipping fleet, and standard shipping orders with private fleet. Hence, supply-side uncertainty in crowdshipping can still result detrimental when is not balanced by the private fleet. Another stream of literature examined the effect of incentives to increase the attractiveness of the delivery task and reduce supply-side uncertainty (Hua et al.,

2020). In this space, research has primarely looked at the role of monetary incentives (e.g., surge pricing) and non-monetary incentives (e.g., work engagement) (Cachon et al., 2017; Qi et al., 2018; Guda & Subramaniana, 2019; Hua et al., 2020; Bernstein, DeCroix, & Keskin, 2021).

However, such solutions are often economically inefficient. For example, Cachon et al., 2017 found that surge pricing is not an optimal solution in terms of platform's profit, and it has raised concerns on the welfare of providers. Indeed, surge pricing tends to increase delivery costs, and makes drivers to compete for such opportunity (Castillo et al., 2022b). Finally, a third stream of literature investigated the key traits of crowdshipping driver behaviors during and post-task. Specifically, Cameron (2022) adopted an exploratory approach to find that crowshipping drivers engage in different behaviors during the service delivery based on their preferences toward being efficient or effective. Castillo et al. (2022b) also explored drivers' attitude toward the service delivery based on the customer's tip. Petriglieri, Ashford, & Wrzesniewski (2019) investigated gig workers commitment toward to the job, finding that lack of organizational identity pushes gig workers to personalize their work identity. Finally, Choudhary et al. (2021) investigated how to motivate drivers to improve their driving performance, and Ta et al. (2021) focused on how framing messages to appeal to drivers enhance participation, performance, as well as post-task satisfaction.

An alternative approach is to examine how delivery task operational characteristics influence drivers' pre-task, task, and post-task behaviors. Operational characteristics of the delivery task comprise of delivery density, location, attended, and expedited delivery (Castillo et al., 2022b). These are important aspects of the delivery task that drivers often consider besides the overall compensation (Castillo et al., 2022b). Usually, upon seeing the announcement, a crowd driver can choose to sign up for it or not, depending on the compensation as well as the opportunity

cost (Fatehi & Wagner, 2022). In this study, we investigate how operational characteristics moderates the relationship between monetary incentives and drivers' behaviors. Specifically, drivers pre-task behaviors refer to the likelihood a driver is to accept the delivery task (Benson, Sojourner, & Umyarov, 2020; Ta et al., 2021). Task behaviors include driver Service time and performance (Liu, He, & Shen, 2021; Castillo et al., 2022b; Fatehi & Wagner, 2022). Finally, post-task behaviors include driver feedback and retention (Liu et al., 2021; Ta et al., 2021). Hence, the overarching research question of this study is RQ: *how do operational characteristics of the delivery task moderate the relationship between compensation and driver behaviors?*

We inform these research questions adopting logic from the emerging empirical literature exploring workers behavior and productivity in service operations, more specifically on the Income Opportunity Effect (IOE) logic developed by (Kamalahmadi, Yu, & Zhou, 2021). This logic explains that in context of just-in-time scheduling, servers positively see work scheduling characteristics as an opportunity to increase their income (Akerlof, 1982). In addition, as an act of reciprocation and goodwill, the servers may exert additional work efforts, especially if it could lead to more opportunities in the future (Spector & Fox, 2002; Avgoustaki & Bessa, 2019). IOE is grounded on Social Exchange Theory (SET) (Thibaut & Kelly, 1959; Lambe, Wittmann, & Spekman, 2001), which predicts that individuals' social behaviors are understood in terms of exchange of resources, because the need to engage in social interactions is driven by the scarcity of resources (Das & Teng, 2002). In this study, we support the logic of IOE with SET. Based on this theoretical background, we develop hypotheses on the effect of delivery task remuneration on driver behaviors in pre-task, task, and post-task. In addition, we hypothesize the moderation effects of delivery task operational characteristics. This study extends IOE and SET by understanding the

impact of task characteristics on explaining behaviors. In addition, while IOE takes the perspective of the employer, we adopt this logic from the perspective of the server.

We empirically investigate these hypotheses by compiling publicly available and proprietary data retrieved from an online retailer using crowdshipping deliveries. The initial dataset includes approximately 6 million delivery tasks with detailed information on driver behaviors and delivery task characteristics over a three-month period. We performed IV/2SLS analysis to decrease the threat of endogeneity biasing the estimations. Overall, we found support for the notion that it is the combination of monetary incentives and operational characteristics of the delivery task (and not just monetary incentives) to affect drivers pre-task, task, and post-task behaviors. This important result contributes to theory and practice. First, we provide empirical evidence for extending SET with IOE in the just-in-time scheduling of crowdshipping, by identifying how task characteristics also influences servers' performance. Second, we also extend SET by unveiling the role of different exchanging partners when drivers accept, perform, and evaluate the exchange. Finally, we extend SET by finding that monetary incentives drive the long-term relationship but not necessarily improve the driver return time. We also offer several managerial contribution relative to mitigating service capacity uncertainty and improving delivery driver performance.

Literature review: Sharing economy challenges in last mile delivery

Recently, the literature has focused on investigating sharing economy business models in last mile delivery (Muñoz & Cohen, 2018). These business models presents characteristics that can potentially disrupt traditional business practices and address last mile delivery challenges (Lim, Jin, & Srai, 2018; Apte & Davis, 2019; Dhanorkara, 2019; Esper et al., 2020; Na, Kweon, & Park, 2021). The sharing economy is a technologically enabled socioeconomic system that grants peers

on-demand and temporary access to underutilized physical and human assets through an online platform (Kathan, Matzler, & Veider, 2016; Eckhardt et al., 2019; Gerwe & Silva, 2020). Sharing economy business models include sharing economy organizations, which are at the intersection of platform organizations and social movements (Gümüsay, 2018).

Sharing economy presents five key characteristics (Eckhardt et al., 2019; Gerwe & Silva, 2020). First, the sharing economy builds on the temporary access to underutilized resources, such as know-how, homes, and labor capacity (Kornberger et al., 2018; Muñoz & Cohen, 2018; Islam et al., 2020; Cui & Davis, 2022). Second, in the sharing economy, there is the transfer of economic value between the peers based on the exchange of critical resources, such as money for goods and services (Filippas, Horton, & Zeckhauser, 2020; Gerwe & Silva, 2020). Third, this exchange is mediated by multisided technology-based online platforms that leverages internet to match and coordinate many small suppliers or service providers to many small buyers (Benjaafar & Hu, 2020; Köbis, Soraperra, & Shalvi, 2021). Hence, data and information transaction are key factors of production in the sharing economy (Davis, 2016; Chen & Wang, 2019; Wang & Wu, 2020). Fourth, consumers cover a major role in sharing economy as they are both providers and users of underutilized resources, becoming prosumers of fully utilized capacity (Eckhardt et al., 2019). Finally, many sharing economy business models (e.g., crowdsourcing) build on a crowdsourced supply, in which the crowd replaces the corporation and provide peers the access to the resources (Kornberger et al., 2018; Lehdonvirta et al., 2019).

Sharing economy business models have also been applied to operations and supply chain management research (Browning, 2020; Hopp & Simchi-Levi, 2021; Hu, 2021), for example to unveil the key challenges of ride-sharing (Hasija, Shen, & Teo, 2020; Mak, 2022), and the impact of crowdsourcing on last mile delivery operations (Carbone et al., 2017; Qi et al., 2018; Fatehi &

Wagner, 2022). Crowdsourcing refers to the sharing economy business model in which logistics tasks once handled and performed by a firm are now assigned to the crowd of individual (Carbone et al., 2017). Crowdsourcing applied to last mile delivery results in crowdshipping (Ciobotaru & Chankov, 2021), which is defined as the practice in which organizations or persons who need to transport a certain items are connecting with individuals of the crowd that are willing perform the deliver (Dayarian & Savelsbergh, 2020). Notable examples of crowdshipping refer to Deliv, Instacart, Amazon Flex, and Walmart GoLocal, which are platforms managing last mile delivery services for retail supply chains (Ta, Esper, & Rossiter, 2018; Garland, 2022). Crowdshipping presents many opportunities to improve last mile delivery. It allows on-demand flexible transportation capacity by leveraging the crowd, for example retail store customers (Devari, Nikolaev, & He, 2017; Dayarian & Savelsbergh, 2020; Howe & Jin, 2022; Richey & Davis-Sramek, 2022). In addition, crowdshipping does not require retailers to invest in fleet assets, because the delivery is performed from an independent contractor who utilizes the personal vehicle (Castillo et al., 2018, 2022a). Crowdshipping also bring the advantage of proximity of fleet capacity because crowdsourced drivers are commonly assigned to a specific geographical area in which they perform deliveries (delivery zone) (Douglas, 2020; Wang & Wu, 2020). This delivery mode presents also characteristics of adaptability to the delivery characteristics and in some instances is a cheaper solution than utilizing a private fleet (Castillo et al., 2018; Tokar & Swink, 2019).

However, the literature has identified a key challenge of crowdshipping related to service provider supply scarcity (Benjaafar & Hu, 2020). In the sharing economy, the supply of service providers is characterized by random capacity, because the total number of available servers in a period of time and in a specific location is random (Dong & Ibrahim, 2020). The platform cannot

exactly predict the number of drivers that will be available to deliver an on-demand order, because delivery drivers independently choose their availability (Carbone et al., 2017). In addition, supply uncertainty and risk increase due competition with other crowdshipping platforms that lowers service providers retention and anti-competitive practices of service providers (Benjaafar & Hu, 2020; Cameron, 2022; Tripathy, Bai, & Heese, 2022). Uncertainty of the pool of service providers is especially important because crowdshipping is often employed for on-demand deliveries that require a fast-shipping option (e.g., 2 hours delivery) (Fatehi & Wagner, 2022). Hence, lacking service providers (hereafter crowdsourced drivers or drivers) ultimately results in an economically inefficient and operationally ineffective delivery. Economically inefficiencies originate from an increase in the delivery cost: to fulfill the order, crowshipping platforms typically improve the attractiveness of the delivery task by increasing the compensation. This reflect the use of surge pricing – the practice of increasing prices during times of high demand to decrease congestion (Cachon et al., 2017; Bernstein et al., 2021). Alternatively, platforms can reduce supply uncertainty by investing resources on the drivers, offering monetary rewards and non-monetary intrinsic motivators (Hua et al., 2020). Crowshipping is often ineffective because supply uncertainty does not guarantee quality in the service provision, resulting in lower on-time delivery rates as compared to a dedicated fleet (Castillo et al., 2018). In crowdshipping, random supply capacity can increase the service queue, which in turn negatively affect waiting-time-sensitive customers outcomes (Ibrahim, 2018; Taylor, 2018).

Therefore, crowdshipping offers a flexible yet challenging resource to use in last mile delivery. Extant research provides some solutions to this challenge. A stream of literature studies the key characteristics and motivations of crowdsourced drivers to shape the usage of delivery drivers. Anderson (2014) and Rosenblat (2016) sorts drivers into three categories, namely

incidental drivers that drive occasionally and are motivated by sharing the vehicle when commuting, part-time drivers that organize their schedule to allocate a consistent number of working hours for the service provider and are motivated by working flexibility, and full-time drivers that deliver full-time and are motivated by income opportunities for underqualified jobs. Cameron (2022) adopts an exploratory approach to find that drivers play either a relational game, in which they engage in positive service encounters in the pursuit of high customer ratings and tips, and the efficiency game, in which drivers aim at minimizing any extra behavior in the pursuit of increasing efficiency and delivery volumes. In last mile delivery, this can be an important discriminant between assigning drivers to attended vs unattended deliveries. Finally, Petriglieri et al. (2019) investigate how the lack of organizational identity pushes gig workers to personalize their work identity. Another stream of literature addresses supplier uncertainty by investigating hybrid fleets composition – combining privately owned delivery vehicles with crowdsourced assets (Castillo et al., 2018). This literature focuses primarily on compensation schemes for private fleet drivers vs full-time vs part-time crowdsourced drivers (Dai & Liu, 2020), and the optimal hybrid fleet design systems to address the cost-service trade-offs by focusing on how delivery costs and order arrival rate intensity relate to compensation mechanisms (Castillo et al., 2022a). Finally, (Castillo et al., 2022b) investigate how customer tipping improves delivery driver last mile delivery performance, finding that this monetary reward can decrease supplier uncertainty. In summary, the literature focuses on attracting and retaining drivers by using almost exclusively compensation mechanisms. However, crowdsourced drivers can maximize their utility also by delivering large volumes (i.e., many orders per batch) as suggested by (Cameron, 2022). Indeed, while delivery drivers may find surge pricing as profitable, they may dislike inconsistency in the offerings (few spotted deliveries for big money vs. a consistent stream of deliveries for low

money). Hence the third research gap refers to identifying the role of delivery volumes in decreasing service providers' supply uncertainty.

Social Exchange Theory and Income Opportunity Effect

Social Exchange Theory

Social exchange theory (SET) is among the most prominent conceptual paradigms for investigating the interactions among individuals and among organizations (Cropanzano & Mitchell, 2005). This theoretical approach was initially developed in purely economic settings to study the interpersonal exchanges, but later its use has been expanded to organizations behavior and business to business marketing relationship (Homans, 1958; Lambe et al., 2001; Das & Teng, 2002). Recently SET has been applied to study the dynamics of buyer-supplier relationships in the supply chain and strategy literature. Studies have addressed the relevance of mutual commitment in the buyer-supplier relationship as a signal of the interest to further develop the relationship (Holm, Eriksson, & Johanson, 1999), the role of power and dependence in the dyadic relationship (Narasimhan et al., 2009), and the reduction of opportunism in the relationship when relational mechanisms are used (Liu, Luo, & Liu, 2009). Finally, studies have further implemented SET as to understand a potential mitigation effect on competition and the partner's performance (Griffith, Harvey, & Lusch, 2006; Terpend & Krause, 2015).

SET asserts that individuals' social behaviors are understood in terms of exchange of resources, because the need to engage in social interactions is driven by the scarcity of resources (Das & Teng, 2002). Following this premise, the exchange is at the core of the interactions between organizations and persons, as they are seeking for rewards or avoiding punishments: a social exchange can be defined as the voluntary action of people motivated by the returns they expect to receive from the interaction (Homans, 1958; Das & Teng, 2002; Griffith et al., 2006). The

exchange is perceive and evaluated on the proportionality between the value that one's behavior provide to the counterpart, and the behavior the counterpart gives back (Homans, 1958). To this extent, the exchange is a reciprocal process that follows the reciprocity rule: one's actions are contingent on the other's behavior (Cropanzano & Mitchell, 2005). Thus, SET interprets organizations or individuals' behaviors as the reward minus the cost of the interaction resulting from the social exchange of resources (Narasimhan et al., 2009).

SET designs the system of social exchange in terms of a series of propositions (Griffith et al., 2006; Narasimhan et al., 2009). First, the success proposition states that members of a relationship will be likely to perform the exchange if the action is rewarded. Second, the reward and value propositions propose that the more valuable the result of the exchange is for an individual, the more likely the member will perform the exchange action again. However, in case the member's action does not receive the expected reward or a punishment, then according to the aggressive proposition the member will avoid the interaction in the future. Fourth, the relationality proposition states that individuals choose the action that maximize the reward considering the probability of receiving a greater reward. Consistently, people are seen as rational and will likely to determine the best possible means to engage in the relationship and maximize the returns (Narasimhan et al., 2009).

Based on these propositions, SET posits that the exchange interactions result in both economic and social outcomes (Thibaut & Kelley, 1959; Lambe et al., 2001). The economic outcomes are strictly related to short-term tangible economic rewards, addressing the parties' financial needs (Cropanzano & Mitchell, 2005). When economic outcomes are used in a social exchange, the parties will focus on the performance to obtain the economic reward, overlooking the long-term relationship (Lambe et al., 2001). Conversely, social outcomes refer to the level of

commitment of the parties and the establishment of a long-term relationship based on trust (Lambe et al., 2001). The parties will develop a long-term orientation in the relationship fulfilling social and esteem needs, and eventually they will consider the long-term benefits of the relationship more valuable than the short-term benefits. In this sense, commitment will function as a signal to the counterpart for the interest in developing the relationship, whereas the trust build on commitment and positive returns will be likely to enhance the long-term relational exchange (Lambe et al., 2001; Cropanzano & Mitchell, 2005).

Finally, parties involved in a social exchange compare the economic and social returns with those offered by alternative parties (Thibaut & Kelley, 1959). Indeed, the parties enter a relationship when the expected returns justify the resource allocated. However, if an alternative party can provide a greater benefit, then the parties involved in the relationship will likely to switch to a different relationship (Lambe et al., 2001). In addition, parties may engage in relational exchanges with different goals and expectations: an organization may be more focused on profit maximization, as compared to another focusing on the relationship development (Lambe et al., 2001). Therefore, the based on the expectations of the parties and the gained returns, parties will either join another relationship or further develop the current one.

Income Opportunity Effect

The IOE refers to the logic grounded on SET that predicts a positive or negative effect of just-in-time scheduling on workers' availability, outcomes, and performance (Kamalahmadi et al., 2021). Based on this logic, a just-in-time scheduling can have a negative effect on worker's performance because workers suffer schedule unpredictability, which leads to a sense of inequity and injustice (Spector & Fox, 2002). However, we contend that the negative effect does not apply to the context of this study, because crowdshipping drivers operate on-demand. Hence, the nature of the task

itself is unpredictable. In contrast, a positive effect occurs when workers perceive the just-in-time scheduling as an income opportunity to work more hours and earn more (Akerlof, 1982). Hence, they are motivated to work and provide a better performance, eventually engaging in act of reciprocation toward the employer, and exerting extra efforts during the service delivery.

Hypotheses development

To formulate the hypotheses of this study, we integrate SET with logic of IOE and literature on crowdsourced drivers. The outcome of interest is driver behaviors, defined at three levels of detail. Driver behaviors pre-task include driver's availability and participation (Ta et al., 2021; Dayarian & Pazour, 2022). Driver task behaviors include driver Service time and the overall performance of the driver (Liu et al., 2021; Castillo et al., 2022b). Finally, driver behaviors post-task include driver satisfaction and retention (Ta et al., 2021; Castillo et al., 2022a). Figure 1 provides an overview of the theoretical model.

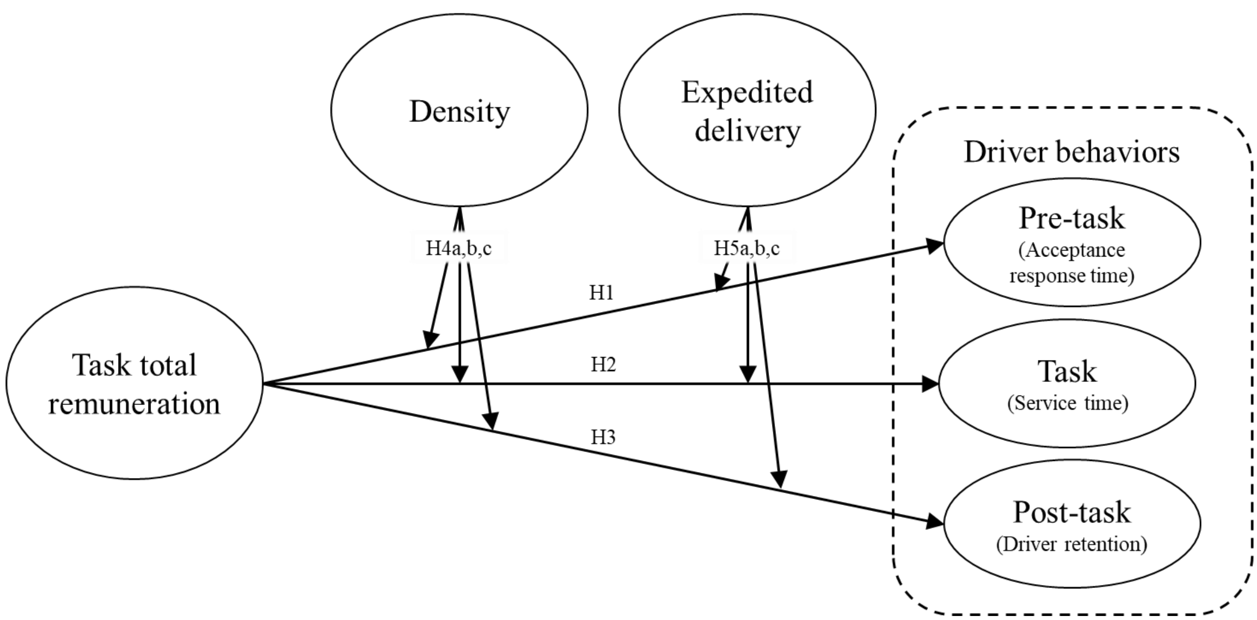


Figure 4 – Theoretical model essay 1

The impact of task remuneration on driver pre-task behaviors

In crowdshipping, task remuneration include two components: compensation and tips (Castillo et al., 2022b). Compensation refers to the fare the platform pays to the driver, whereas tips are consumers' voluntary payments of money to the service provider (Alexander, Boone, & Lynn, 2021).

On-demand platforms broadcast a delivery task associated with the remuneration. As an act of exchange, they voluntary commit to the task and complete the delivery motivated by the return expected form the interaction (Das & Teng, 2002). Hence, according to the first step of SET, the reward (i.e., remuneration) will drive the service provider's pre-task behavior to accept the delivery task. Higher is the reward, more attractive will be the delivery task. An extensive body of literature has investigated how monetary rewards drive crowdshipping drivers acceptance rate (Cachon et al., 2017; Qi et al., 2018; Hu, 2021; Castillo et al., 2022b). As a consequence, the time taken for a delivery task to be accepted once broadcasted in the system decreases as the total remuneration for the task increases. Thus, we expect that:

Hypothesis H1: *Delivery task remuneration is positively associated with driver pre-task behaviors on acceptance response time. A higher remuneration will decrease the time elapsed between the delivery task is broadcasted and accepted.*

The impact of task remuneration on Driver task behaviors

Upon accepting the delivery task, the driver commute to the pick-up point (i.e., the store) to collect all the parcel units to be delivered. In crowdshipping, the Service time typically denotes the amount of time it takes to a driver to deliver the order, including the travelling time and the on-site Service time (the time for parking, finding the apartment, and deliver the package) (Liu et al., 2021; Fatehi

& Wagner, 2022). Service time also bears an important consequence relative to the driver delivery performance (i.e., on-time delivery) (Castillo et al., 2022b).

Following SET, a higher reward resulting from the exchange will motivate parties to commit to the relationship and meet the expectations of the counterpart (Cropanzano & Mitchell, 2005). IOE adds that servers often work harder to increase the ratio of remuneration per hour (Kamalahmadi et al., 2021). Hence, crowdshipping drivers will likely decrease their Service time and increase delivery performance upon receiving a higher compensation because servers reciprocate the employer by committing extra effort to provide superior customer service (Flynn, 2005). This prediction is also rooted in the human resource literature: remuneration is perceived as the recognition of the service delivery performance, and servers who earn more typically exert an effort to improve performance (Jerkings Jr. et al., 1998; Mitchell & Mickel, 1999; Liao et al., 2009; Cerasoli, Nicklin, & Ford, 2014). Thus, drivers receiving a higher compensation will improve their performance through behaviors that reduce the Service time and improve performance, such as following the planned delivery sequences from the routing tools (Liu et al., 2021), and reducing their errors during the delivery task (Awaysheh et al., 2021). Hence, a higher delivery task remuneration will positively impact Driver task behaviors by reducing the Service time. Hence, we expect that:

Hypothesis H2: *Delivery task remuneration is positively associated with Driver task behaviors on Service time. A higher remuneration will decrease the Service time.*

The impact of task remuneration on driver post-task behaviors

Driver retention is indeed a central element in crowdshipping. Retention is defined as the driver future intentions and availability to perform a delivery task (Cantor, Macdonald, & Crum, 2011; Dayarian & Pazour, 2022). SET predicts that satisfaction among the parties results in a long-term

relationship of commitment and trust, characterized by relational norms, defined as the expectation that the counterpart will engage in behaviors based on mutually agreed upon rules (Lambe et al., 2001; Cropanzano & Mitchell, 2005). IOE elaborates this prediction in the service industry, by suggesting that workers seeing a positive effect on just-in-time scheduling will commit to provide service performance in the future because it leads to more income opportunities (Kamalahmadi et al., 2021). In the crowdshipping context, driver retention has a crucial operational aspect because it reduces supply-side uncertainty. Therefore, a higher remuneration of the delivery task increases driver post-task behaviors retention.

Hypothesis H3: *Delivery task remuneration is positively associated with driver post-task behaviors on satisfaction and retention. A higher remuneration will increase satisfaction and retention.*

The moderating role of delivery task operational characteristics

Delivery task operational characteristics include route drop-offs density and delivery type. We chose these two operational aspects of the delivery task because of their meaningful value for driver behaviors pre-task, task, and post-task.

In last mile delivery, density is defined as the number of miles a driver must travel to deliver each order in the route (Wang, Rabinovich, & Guda, 2022). Density has been used in prior literature as a key factor influencing delivery efficiency (Boyer, Prud'homme, & Chung, 2009; Castillo et al., 2022b), and a core element to achieve a competitive advantage (Zhang et al., 2018; Wang et al., 2022). Prior literature founds that density in drop-off locations affect the Service time and delivery operations (Liu et al., 2021), because it enhances the economies of scale that could be obtained by delivery operations focused on a specific areas (Boyer et al., 2009; Wang et al., 2022). In other words, at higher levels of density, delivery operations will be concentrated in a

specific geographical areas (i.e., shorter route to perform), and greater effectiveness in deliveries is achieved, whereas lower levels of density will result in more geographically dispersed drop-off points that will decrease the economies of scale effect on delivery performance.

Crowdshipping drivers want to maximize the remuneration per hour (Cameron, 2022). This translates into the number of drop-offs that can be completed in a specific amount of time. Hence, delivery tasks that concentrate the drop-offs in a specific area will be more attractive to delivery drivers, who will see the opportunity to complete the task with greater efficiency in the delivery costs (e.g., gas), and greater chances to provide an on-time performance. In addition, drop-off density will reduce the driver Service time and on-time performance. For example, drivers can find density convenient to reduce the on-site Service time (e.g., same parking lot, same apartment building). Finally, using notions from SET and IOE, density increases the reward associated with the delivery, increasing satisfaction, and motivates drivers to engage in future delivery tasks, increasing retention. Therefore, drop-offs density enhances the relationships between remuneration and driver behaviors. We expect that:

Hypothesis H4a,b,c: *Drop-offs density moderates the relationship between delivery task remuneration and driver behaviors, that is higher drop-offs density enhances the positive relationships between remuneration and a) driver pre-task, b) task, c) post-task behaviors.*

The second moderator refers to the type of delivery task. IOE discusses a difference between short-notice and real-time notice (Kamalahmadi et al., 2021). The discriminant is the amount of time a server is noticed about the task. While both are typical in crowshipping, a key difference is that with short-notice, a driver has time to plan ahead the delivery provision, whereas with real-time notice, the driver is requested to accept and complete the delivery task in a shorter amount of time, providing little opportunity to adjust their other plans. In the context of this study,

drivers can accept delivery task including expedited or standard deliveries. A real-time notice is associated with expedited deliveries, which are typically requested by customers for urgent deliveries. Standard and regular deliveries, in contrast, correspond to short-notice, because drivers are given the opportunity to plan ahead for the delivery task.

We argue that expedited deliveries are less preferrable to drivers. First, despite the income opportunity, drivers need to adjust their plans after accepting a last-minute delivery task, reducing the flexibility related to crowdshipping, a core motivational factor for drivers (Anderson, Allen, & Browne, 2005; Rosenblat, 2016; Benjaafar & Hu, 2020). Second, drivers also account for the feasibility of the delivery task. In these instances, the driver accepting an expedited delivery task is bounded to completing the task in a short time frame, which can lead to a delay and lowering the performance of the delivery driver. Indeed, crowshipping platforms often measure driver performance based on the on-time arrivals (see for example Walmart spark driver evaluation (Walmart.com, 2023)). An expedited delivery indicates a lower amount of time to complete the task, thus a greater likelihood of a late delivery. Relatedly, the driver task behaviors on delivery performance reduces, along with the Service time. As a result, driver post-task behaviors are also affected by the delivery type. For an expedited delivery, the driver had to work harder, decreasing the satisfaction and retention. We expect that:

Hypothesis H5a,b,c: *Delivery type moderates the relationship between delivery task remuneration and driver behaviors such that, expedited delivery mitigates the positive relationships between remuneration and driver a) pre-, b) post-task behaviors, and delivery performance, while (c) expedited delivery enhances the positive relationships between remuneration and Service time.*

Empirical Setting and Data

Data description

We empirically investigated these hypotheses by compiling a dataset from multiple sources. First, we retrieved delivery operational data and driver outcomes data from a Fortune 100 retailer (hereafter called *Alpha*). The retailer has launched its white-label crowdshipping platform, which performs home deliveries from the retailer's stores using crowdsourced drivers. Upon being pinged from the retailer for home deliveries, the crowdshipping platform broadcasts the offer for the delivery task through the Driver App, where the offer is visible to a set of drivers assigned to a driver zone[1]. The offer includes details related to the compensation, the pre-tipping, the delivery type, the number of orders included in the delivery task, the total of miles that a driver would drive to perform the task (i.e., from the store to the final drop-off), as well as the store and customers' address. Upon accepting the task, the Driver App provides GPS navigation instructions to arrive at the pick-up point and drop-off the orders at the customer's destination.

Alpha shared a raw dataset covering three months (February to April 2022), and containing approximately ~4 million completed (i.e., delivered to final customer) delivery tasks associated with approximately ~7 million customers' orders. The dataset includes detailed information about the delivery tasks and the drivers, and reports several date and time stamps reporting the hour, minute, and second related to the flow of a delivery task i. Initial offer sent ($inos_i$) refers to the timestamp of when the offer was initially broadcasted to the crowdsourced drivers within the delivery assigned zone. Offer acceptance ($ofac_i$) refers to the timestamp of when the driver accepted the delivery task. Driver arrival ($drar_i$) refers to the timestamp of when the driver arrived

[1] As common in crowdshipping (Guda & Subramaniana, 2019; Tripathy et al., 2022), the platform assigns drivers to a driver zone, which serves the stores located in at least one zip code (i.e., the same zone may serve multiple zip codes).

at the pick-up location. Driver starts trip ($drst_i$) refers to the timestamp of when the driver started the delivery route relative to a specific drop-off within the delivery task. Driver arrival at destination ($arad_j$) refers to the timestamp of when the driver arrives at the drop-off destination. Order delivered ($orde_j$) refers to the timestamp of when the driver delivers the order. Finally, delivery time window end (de_j) refers to the planned delivery time for a specific order. This dataset also includes the delivery type for each order (expedited, standard, regular), and the intrinsic characteristics of each order in the delivery task, for example volume and weight. Further, the dataset reports the remuneration for the delivery task, distinguishing between compensation and pre-tipping, and driver outcomes. Finally, the dataset reports the geo coordinates (latitude and longitude) of each drop-off location, as well as information relative to the store that dispatched the delivery task, including store id and complete address.

Second, we manually retrieved publicly available data from *Alpha*'s website about stores that dispatched a delivery task. Specifically, we downloaded data on stores' id, type, and exact address. *Alpha* sorts stores dispatching a delivery task into different[2] types, based on the dimension of the store. The exact address includes street name and number, 5-digit zip code, town, and state. Next, we built a custom program to extract the geographical coordinates (i.e., longitude and latitude) from each store address. Following current literature (Belo, Ferreira, & Telang, 2014; Belenzon, Chatterji, & Daley, 2020; Barrios, Hochberg, & Yi, 2022), we employed Google Maps Geocoding application programming interface (API), which is a Google Cloud Platform powered by Google Enterprise API, that converts between addresses and geographic coordinates. The platform library provides an XML file to automate the search for coordinates from the specified address and guarantees complete integration with Microsoft Excel. Among the advantages of this

[2] For confidentiality reasons, the number of drivers, stores, zip codes, delivery zones, and store types are not disclosed.

API is the correction of small variations in spelling (misspelling) between the manually retrieved and the actual address, allowing to find the geocoordinates when there is a close match (Belenzon et al., 2020). We manually checked a sample of these conversions to ensure the precision of the conversion, without finding any unreasonable matching (Belenzon et al., 2020). Finally, we matched this and *Alpha* datasets through store id. Hence, using the geocoordinates of drop-off locations (from *Alpha*) and of store locations (from Google), we computed the store-drop-off distance, expedited in miles, using *geodist* function in STATA17. This function computes ellipsoidal distances (i.e., "the length of the shortest curve between two points along the surface of the mathematical model of the earth WGS 1984 datum" (Picard, 2022) – the same used by Google Earth) using Vincenty (1975)'s equation (Picard, 2022). A limitation is that geodist computes the actual distance, not the travel distance. However, the nature of the dataset being limited to local store-to-home deliveries allows to reasonably assume that g_n and travel distance are similar and strongly correlated.

Third, we retrieved publicly available data from the Internal Revenue Service's (IRS's) Statistics of Income (SOI) database relative to the per capita income at zip code level. The SOI database reports several items forming the total income for a zip code in a specific year, such as the adjusted gross income (AGI), income from royalties, income form wage and salary, and income from unemployment insurance benefits. IRS compiles this database from tax returns filed by each zip code in each year. Similar to prior literature (Mulvey, 1983; Cunningham, Gerardi, & Shen, 2021), for each zip code and latest available year (2019), we retrieved the per capita income from the total income amount and the number of returns with total income. Finally, we matched this dataset with the other two by zip codes.

Data cleaning

Before performing the main analyses, we cleaned the dataset following best practices of recent literature investigating similar contexts (Farber, 2015; Miao et al., 2022). First, we removed incomplete or erroneous observations and outliers for all the variables of interest. Specifically, we removed (1) orders with missing timestamps, (2) orders delivered by drivers whose age was below 21 and above 71[3], (3) erroneous observations with total travel distance longer than 21 miles[4] (4) erroneous observations of ordered delivered before 6 AM and after 10 PM[5]. In the end, 1,667,993 million observations were retained as final sample. Based on the cleaned dataset, we plot the histogram for the delivery tasks across the calendar days of the dataset in Figure 5, across the days of the week in Figure 6, and the clock hour of the days in Figure 7. These histograms reveal a positive trend of number of delivery task completed across the three months, with the pick during mid-April, a reasonably consistent demand of delivery task across the days of the week and the clock hour of the days, with peaks early in the morning (7-8 AM) and in the afternoon (1-2 PM). To ensure that such trends would not be a source of endogeneity, we introduced dummy controls for month, day of the week, and time of the day in all estimations.

[3] These thresholds result from symmetrically winsorizing driver age at 99%, align with prior literature investigating similar contexts (i.e., crowdshipping) (Ta et al., 2021), and are motivated by federal regulations on alcohol delivery, which may create unnecessary noise in the dataset (e.g., a 18 years old driver may not be given the choice of a delivery task of orders including alcohol delivery).

[4] This threshold result from symmetrically winsorized distance traveled at 99%, which produces an average distance traveled of 4.3 miles (SD = 4). This aligns with prior literature investigating similar context (i.e., crowdshipping), for example, Miao et al. (2022) average trip distance was 11km (~7 miles), and Castillo et al. (2022b)'s netnography report drivers performing between 7 and 10 miles.

[5] *Alpha* provides the delivery service from 7 AM to 9 PM. Thus, a driver can accept a delivery task between the 6 AM hour and no after the 9 PM hour of a given day.

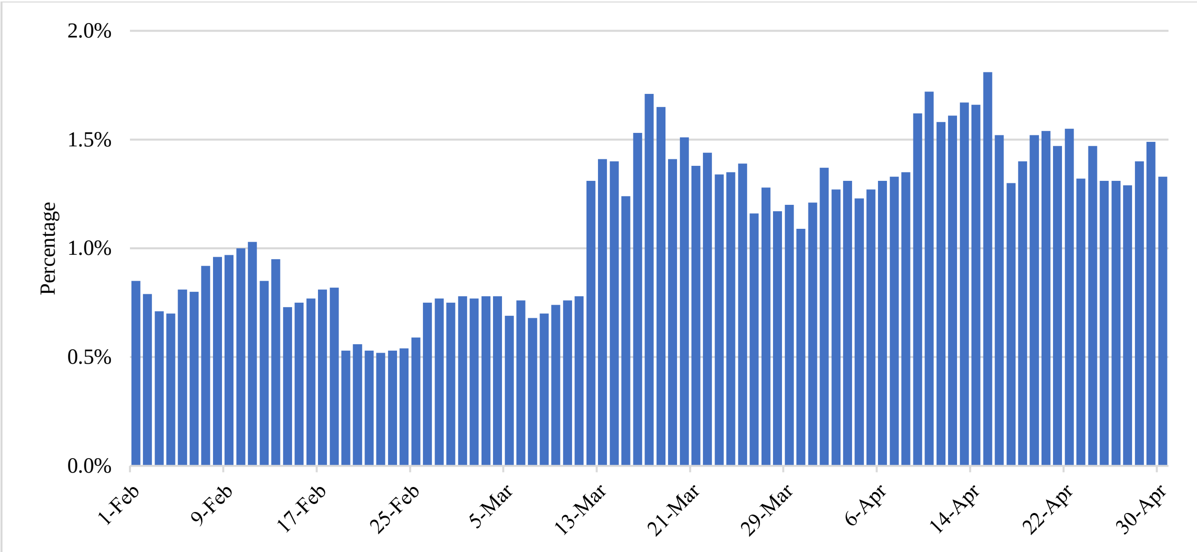


Figure 5 – Histogram for delivery tasks across calendar days

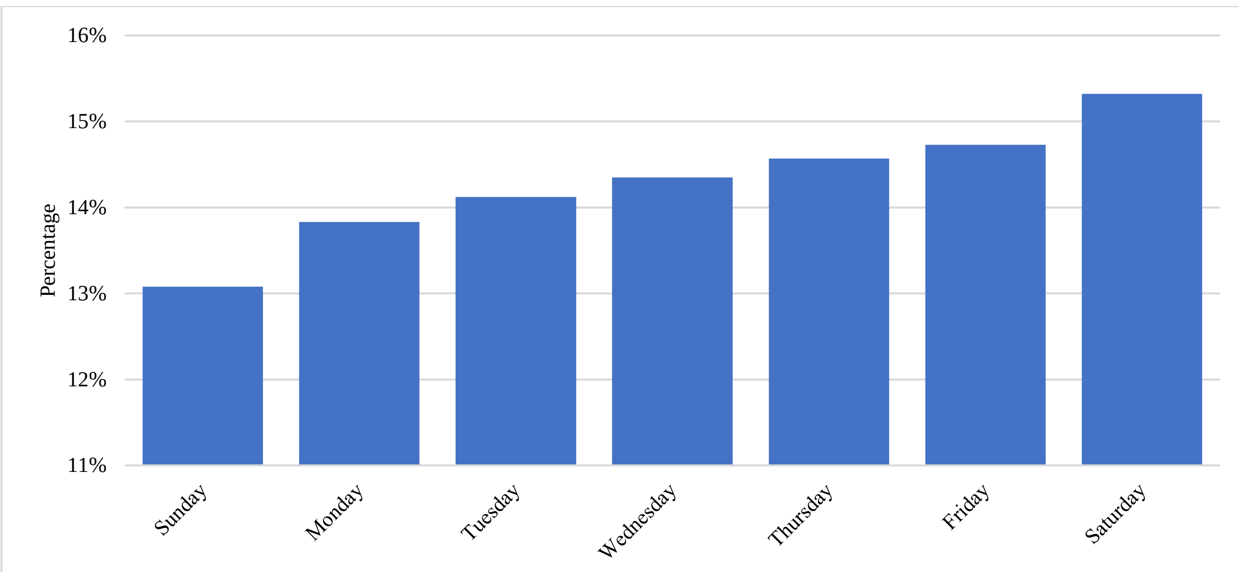


Figure 6 – Histogram for delivery tasks across days of the week

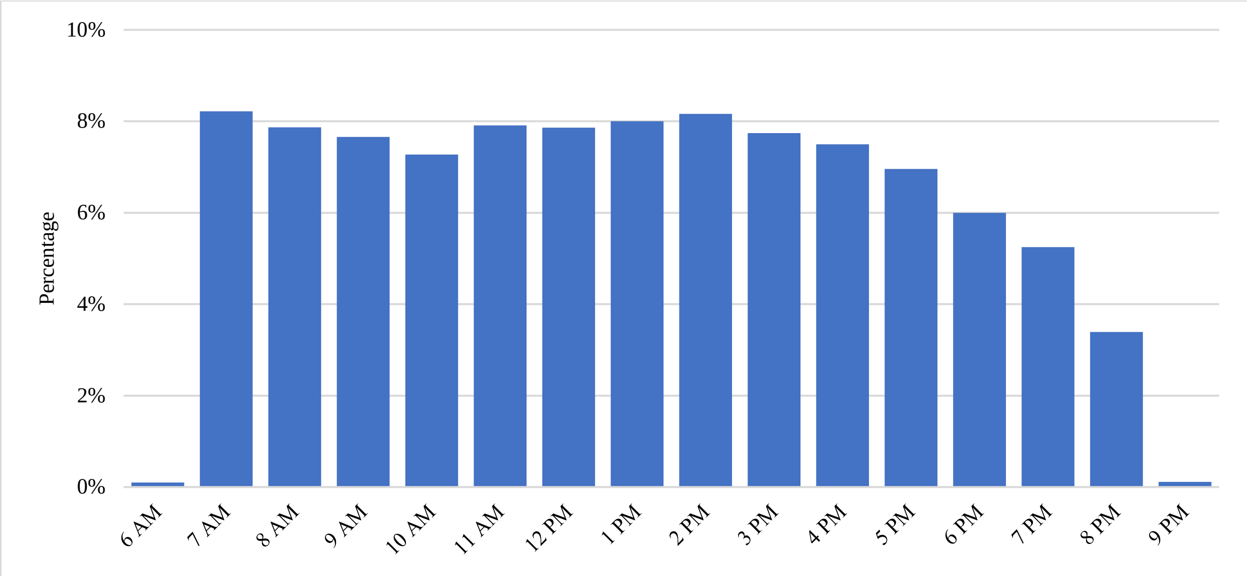


Figure 7 – Histogram for delivery tasks across clock hours

Variables construction - Outcome variables

We computed the outcome variables and predictors following theory, prior literature, and best practices. Table 1 reports the descriptive statistics, and Table 2 reports the correlations.

Table 1 – Descriptive Statistics

	Variable		Mean	SD	Min	Max	Source
(1)	Acceptance (minutes)*	A_i	6.38	11.74	0.05	58.74	(Benson et al., 2020)
(2)	Service time (minutes)*	ST_i	15.19	7.99	3.04	42.7	(Liu et al., 2021)
(3)	Driver attrition	$DRAT_d$	0.11	0.32	0	1	(Castillo et al., 2022b)
(4)	Driver return time	$DRET_d$	2.77	2.81	1	88	(Song, 2022)
(5)	Remuneration ($)*	M_i	27.81	11.82	6.43	94.79	(Castillo et al., 2022b)
(6)	Density (order/miles)*	h_i	0.87	0.72	0.16	4.19	(Wang et al., 2022)
(7)	Expedited	E_i	0.1	0.3	0	1	(Peinkofer, 2020)
(8)	Cumulative number delivery task	cum_{fd}	2.45	1.65	1	21	(Ergün-Şahin et al., 2022)
(9)	Driver experience	exp_d	35.99	25.28	0	88	(Batt & Gallino, 2019)
(10)	Distance drop-off to store new pickup (miles)	$dist_i$	0.02	0.17	0	1.63	(Liu et al., 2021)
(11)	Driver age (years)*	age_d	41.86	10.77	21	70	(Ai et al., 2023)
(12)	Number drivers in zone	ndr_z	12.84	10.38	1	92	(Tripathy et al., 2022)
(13)	Number orders in batch*	nor_i	1.75	0.52	1	12	(Lim, 2023)
(14)	Number of unattended deliveries*	nun_i	1.14	0.55	0	7	

Note: * The variable was symmetrically winsorized at 99%. i denotes the delivery tak. d denotes the driver id. f denotes the calendar day. z denotes the delivery zone.

Table 2 – Correlations

	Variable	(1)	(2)	(3)	(4)	(5)	(6)
(1)	Acceptance (minutes)	1.00					
(2)	Service time (minutes)	0.05***	1.00				
(3)	Driver attrition	0.00***	0.01***	1.00			
(4)	Driver return time	-0.00*	0.00*	0.10***	1.00		
(5)	Remuneration ($)	-0.06***	0.20***	-0.08***	-0.10***	1.00	
(6)	Density (order/miles)	-0.04***	-0.37***	0.01***	0.02***	-0.10***	1.00
(7)	Expedited	-0.05***	-0.11***	-0.02***	-0.01***	-0.10***	-0.11***
(8)	Cumulative number delivery task	-0.00***	-0.04***	-0.08***	-0.10***	-0.02***	0.03***
(9)	Driver experience	-0.01***	-0.05***	-0.09***	0.10***	-0.01***	0.03***
(10)	Distance drop-off to store new pickup (miles)	0.00	0.00***	-0.04***	-0.06***	0.00*	-0.00
(11)	Driver age (years)	-0.04***	0.02***	-0.02***	0.00	-0.00	-0.00***
(12)	Number drivers in zone	-0.07***	-0.05***	-0.05***	-0.05***	-0.05***	0.13***
(13)	Number orders in batch	0.03***	0.17***	-0.07***	-0.08***	0.09***	0.17***
(14)	Number of unattended deliveries	-0.00*	0.42***	-0.09***	-0.09***	0.17***	-0.12***

	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)
(7)	1.00							
(8)	-0.00	1.00						
(9)	0.02***	0.06***	1.00					
(10)	-0.00**	-0.00	0.01***	1.00				
(11)	-0.01***	0.04***	0.07***	0.00	1.00			
(12)	0.04***	0.09***	0.14***	0.00	0.02***	1.00		
(13)	-0.49***	0.01***	-0.00***	0.00	0.01***	0.01***	1.00	
(14)	-0.15***	0.01***	0.00***	0.00*	0.01***	0.00***	0.35***	1.00

Note: * p < .1 ** p < .05 *** p < .01

Delivery driver pre-task behavior, Acceptance response time (A_i), was operationalized as the elapsed time (in minutes) between Initial offer sent ($inos_i$) and Offer acceptance ($ofac_i$). For the delivery task i, Acceptance response time was computed as follows:

$$A_i = ofac_i - inos_i$$

Larger values of $Acceptance_i$ indicates that the delivery task took longer to be accepted. Unfortunately, for confidentiality reasons, the dataset does not report the time stamp of when the delivery driver entered the Driver App on a given day. Hence, for the first trip of each day, we assumed the driver who accepted the delivery task was active (i.e., had his/her phone available) at the time the offer was broadcasted. Despite this limitation, similar to Yu et al. (2020), we support the validity of this operationalization using prior literature and theory. Similar operationalization has been used in the studies related to gig economy workers (Benson et al., 2020). In addition, Hu (2021) discussed how crowdshipping drivers face a setup cost when start delivering on a specific day, as they stay around not just for one ride-task but for many. Thus, the accepting driver intended to perform a series of delivery task and planned ahead before start delivering. Prior to the final analysis, we computed the natural logarithm for A_i to correct for distribution skewness. A similar approach had been used in prior literature (Andritsos & Tang, 2014; Chan et al., 2021).

Delivery Driver task behavior, *Driver Service time* (ST_i), was operationalized as the sum of the elapsed time (in minutes) between Driver starts trip ($drst_{ij}$) and Driver arrival at destination ($arad_{ij}$) for each order. For the delivery task i comprising of j number of drop-offs, *Driver Service time* (ST_i) was computed as follows:

$$ST_i = \sum_{j=1}^{J} \left(arad_{ij} - drst_{ij}\right)$$

Larger values of ST_i indicates that the assigned driver took longer to complete all the deliveries in the task. Similar operationalizations have been used in the literature to indicate the amount of Service time for a delivery driver (Liu et al., 2021). Driver Service time ranges between 5 and approximately 50 minutes. This aligns with prior literature reporting, for example, that drivers of

Amazon logistics partners average between 80/200 stops per shift (i.e., for 80 stops in a 8-hour shift, a driver would take 6 minutes/drop-off) (Wang et al., 2022), and industry benchmarks, for example, Amazon delivery partners usually have 170 to 350 packages per delivery shift (Mayo, 2021), and UPS delivery drivers make at least 100 packages in a 9-11 hours shift (Straightaway, 2022).

Finally, we computed Driver retention using two distinct operationalizations. The first is *Driver attrition (DRAT)*, measured as a dummy variable taking 1 when a driver quits (i.e., disappears from the dataset), 0 otherwise. This operationalization follows prior literature investigating individuals' return behavior (De Vries, Roy, & De Koster, 2018), investigating crowdshipping delivery drivers (Cullen & Farronato, 2021; Castillo et al., 2022b), and is of great interest to firms as it is a costly and operational challenge (Emadi & Staats, 2020). The second is *Driver return time (DRET)* after the delivery task, computed as the number of days until another delivery task is completed or the end of the dataset. This operationalization follows prior literature investigating the return time of individuals (De Vries et al., 2018; Song et al., 2022). A common limitation of these operationalizations refers to the limited time span of the dataset, which does not allow to observe the actual time of the next delivery task for all the delivery tasks, specifically those accepted close to the end of the dataset. In these instances, common practice suggests computing the number of days until the next delivery task or the end of the dataset (De Vries et al., 2018). This operationalization is required to account for the censored nature of the dataset (Hosmer & Lemeshow, 1999). A greater number of days between delivery tasks denotes lower retention. The dataset reveals that drivers return on average every 3 days, and *DRET* ranges between 1 and 88 days.

Variables construction – Focal predictors

The focal predictors were operationalized following prior literature (Castillo et al., 2022b). Specifically, *Delivery task remuneration* is computed as the sum (in USD) of the compensation and tips for each order in the delivery task. For the delivery task i comprising of j number of drop-offs, *Delivery task remuneration* (M_i) was computed as follows:

$$M_i = \sum_{j=1}^{J} m_j$$

Where m_j is the remuneration for the single drop-off j. Delivery task remuneration ranges between \$6.5 and \$103. Other crowdshipping platforms, such as Instacart, reports instances of remuneration as low as \$5 and average \$24 per batch (i.e., delivery task) (Rhea, 2022).

The moderator density was computed following Wang et al. (2022). Density is the ratio between the total travel distance (in miles) to complete the delivery task, and the number of drop-offs in the delivery task. For the delivery task i, *Density* (h_i) was computed as follows:

$$h_i = \frac{V_i}{\sum_{n=1}^{N}(g_n)}$$

Where g_n is the distance in miles to reach delivery stop n from the previous delivery stop in the route. The starting point is always the store. Each following delivery stop corresponds to each delivery drop-off. Higher values of h_i indicates greater density, as a higher number of drop-offs per mile. The second moderator, *Delivery type* (*DT*), was operationalized as a dummy variable indicating whether the delivery task was expedited or standard (Peinkofer et al., 2020).

Variables construction – Control variables

We also included a set of control variables. First, we included control variables relative to the delivery driver who performed the task. We accounted for the cumulative number of deliveries that a driver performed on a given day before performing the focal one. This variable captures the

level of fatigue of a delivery driver throughout the day. Prior literature identified fatigue as a main predictor of workers performance throughout the day (Huffman & Bognanno, 2018), and workers' selection of easy tasks (Diwas et al., 2020). This operationalization follows operations management literature that capture fatigue as, for example, the cumulative number of patients served (KC & Terwiesch, 2009; Ergün-Şahin et al., 2022). Then, we accounted for the driver experience, captured by the number of times a driver delivered prior to completing the focal delivery task (Mao et al., 2019). Despite the major limitation of this control variable relative to the censored nature of the dataset, we followed prior literature in capturing the effect of the cumulative delivery experience (Batt & Gallino, 2019). Next, we included the distance (in miles) between the last drop-off location of the focal delivery task and the store of the new pick-up. This variable takes 0 when the focal delivery task was the first of the day, because the dataset does not capture the location of the driver before accepting the first task, and when the delivery task was the last of the day. We included this variable as a proxy to control for the time it takes for a driver to return to the dispatching center (store), following prior literature (Liu et al., 2021). Finally, we capture driver age as the sole demographic variable that we could access. Recent literature investigated the key importance of age for driver commitment to the ride-sharing platform (Ai et al., 2023).

Second, we included the number of drivers that were active in the same delivery zone of the driver who performed the delivery task, at the clock time (i.e., same hour) when the delivery task was broadcasted (Miao et al., 2022). Despite in the context of this study delivery drivers do not know how many other drivers perform deliveries at a given time and day, we included this control variables to exclude potential variance explained by drivers collusion, surge pricing, and more in general delivery supply uncertainty (Benjaafar & Hu, 2020; Banerjee, Freund, & Lykouris, 2022; Miao et al., 2022; Tripathy et al., 2022).

Finally, we included control variables relative to the deliver task, including the number of orders to be dropped-off, and the number of unattended deliveries, which impact the time it takes to perform the delivery task (Campbell & Savelsbergh, 2005; Wang et al., 2016; Agatz, Fan, & Stam, 2021). In addition, we accounted for fixed effects related to store type, because prior literature identified the importance of stores in order fulfillment (Gao & Su, 2017; Ishfaq & Raja, 2018; Dayarian & Pazour, 2022), and time fixed effects related to months, days of the week, and time of the day (Drake et al., 2020; Choudhary et al., 2021; Salari, Liu, & Shen, 2022). Time of the day was captured dividing the day into morning (6:00 AM – 12:59 PM), afternoon (1:00 PM – 4:59 PM), and after hours (5:00 PM – 9:00 PM) per (Drake et al., 2020).

Empirical Models

Preliminary analysis

Prior to the main analysis, we symmetrically winsorized the outcome variables, focal predictors, and control variables at 99% to limit the effects of outliers and nonlinearities (Dikolli et al., 2021; Qi et al., 2022). In addition, despite the correlation table revealed potential source of multicollinearity, we observed that all Variance Inflation Factors are in line with literature commonly accepted thresholds (Kennedy, 2008; Lumineau & Henderson, 2012; Perdikaki, Peng, & Heim, 2015). Finally, following best practices (Aiken, West, & Reno, 1991) and common procedures in the operations management literature (e.g., Kim & Zhu, 2018; Amengual & Apfelbaum, 2021; Delfgaauw et al., 2022), in estimating the effect of interactions between continuous independent variables, prior to the empirical analysis, we mean-centered the focal predictors.

Endogeneity concerns

The purpose of this study is to investigate the impact of monetary and operational characteristics of a delivery task on driver behaviors. Despite the effort to collect many observable variables, upon consulting operations management literature, we identified potential sources of endogeneity. Endogeneity is an empirical challenge that can manifest through reverse causality bias, sample bias, omitted variable bias, estimation model bias (Ho et al., 2017; Lu et al., 2018; Mithas et al., 2022). We do not expect to face endogeneity for reverse causality bias, because the variables of interest follow the *flow* of the delivery task (e.g., drivers see remuneration before accepting the task, and task acceptance will not change the remuneration disclosed on the Driver App), and time stamps were collected in real-time. Prior operations management literature investigating similar research questions confirms that in this setting, reverse causality is not a concern (Wang & Zhou, 2018; Patel et al., 2021; Jain & Tan, 2022). However, in the analyses of this study, we addressed three potential sources of endogeneity, namely endogeneity for sample bias, omitted variable bias, and model estimation bias.

Endogeneity for sample bias (or sample selection) occurs when a dependent variable missing values are the result of a non-random process that affects the OLS estimations given a missing biasing factor influencing the presence of the observation (Certo et al., 2016). As discussed later in this section, in the first stage of the 2SLS approach, we regress remuneration on the control variables and the instrumental variables. However, the final sample included observations for 2.5 million observations for remuneration. Hence, we consider whether the estimated effects of such *smaller* sample might be systematically biased due to unobservable factors and want to assess that a smaller sample is systematically biased by observable variables in the dataset. Following prior operations management literature (Lu & Shang, 2017; Wei, Xiao, & Rong, 2021; Barker et al.,

2022), to avoid this source of endogeneity, we adopt Heckman's sample selection model (Heckman, 1979). Specifically, we calculated the inverse-Mills ratio and included it the main IV/2SLS regressions to control for the potential selection bias (Dixon, Hong, & Wu, 2021; Wei et al., 2021). We estimated the inverse Mills ratio through a probit regression predicting the occurrence of observing remuneration in the dataset (Gambeta, Koka, & Hoskisson, 2019). Hence, similar to Wei et al. (2021), we computed the probability of a delivery task disclosing remuneration for the full sample (~ 6 million delivery tasks) using the moderators, control variables, the average per capita income, and additional fixed effects relative to the driver delivery zones fixed effects (as a robustness check, we estimated the inverse Mills ratio replacing delivery zone with zip codes fixed effects, finding consistent estimations). We include additional control variables and fixed effects that are not included in the outcome equation for the exclusion restriction (Lu & Shang, 2017). Following prior literature (Lu & Shang, 2017; Shang, Ferguson, & Galbreth, 2019; Suk, Lee, & Kross, 2021), we performed the probit regression on a binary dependent variable (Remuneration available 1 = yes 0 otherwise) with the following selection equation:

(1) $$Remuneration_i = \begin{cases} 1 & if\ Z_i\Gamma + \varepsilon_{id} > 0, \\ 0 & otherwise \end{cases}$$

Where Z_i is the vector of independent variables, Γ is the vector of coefficients, and ε_{id} are robust standard errors, clustered for driver id to mitigate the potential source of heteroscedasticity. $Z_i\Gamma$ is specified as follows:

(2) $$Z_i = \gamma_0 + \gamma_1 h_i + \gamma_2 E_i + \gamma_3 DC_i + \gamma_4 cum_d + \gamma_5 exp_d + \gamma_6 dist_i + \gamma_7 age_d + \gamma_8 ndr_z + \gamma_9 nor_i + \gamma_9 nun_i + \gamma_{10} pci_p + \gamma_{11is} store_{is} + \gamma_{12im} month_{im} + \gamma_{13iw} dow_{iw} + \gamma_{14it} time_{it} + + \gamma_{15iz} zone_{iz}$$

Where pci_p is the per capita income at p zip code level retrieved from IRS 2019, $store_{is}$, $month_{im}$, dow_{iw}, $time_{it}$, $zone_{iz}$, are the categorical dummy variables for, respectively, store type

s, month m, day of the week w, time of the day t, and delivery zone z. Table 3 reports the result of equation (1). From equation (1), we obtain the inverse Mills ratio (IMR) and include it in all regression models in this study.

Table 3 – Heckman's model first stage

DV: 1 = Remuneration available, 0 = otherwise		
h_i	0.045****	(0.001)
E_i	-0.108****	(0.002)
DC_i	-0.000****	(0.000)
cum_{fd}	0.030****	(0.000)
exp_d	-0.004****	(0.000)
$dist_i$	-0.000****	(0.000)
age_d	-0.001****	(0.000)
ndr_z	0.001****	(0.000)
nor_i	-0.184****	(0.001)
nun_i	-0.036****	(0.001)
pci_p	0.001****	(0.000)
Constant	-0.060****	(0.017)
Month fe	YES	
Day of the week fe	YES	
Time of the day fe	YES	
Delivery zone fe	YES	
χ^2	861094.669	
N	6,120,444	

Note: $^{+}$ $p < 0.1$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.005$, **** $p < 0.001$. Reported standard errors are clustered by driver ID.

Next, we addressed endogeneity for omitted variable bias, which manifests when unobservable or unavailable factors affect both predictors and the outcome variables (Ho et al., 2017). The impact of the focal predictor, delivery task remuneration, on the outcome variables, are biased for distinct reasons. Prior literature reports several factors explaining the impact of delivery task compensation on driver behaviors. First, predicted delivery service requests (predicted demand) and delivery capacity (driver supply) are core elements impacting the remuneration for a delivery task, that with the prior literature studying phenomenon such as surge pricing, collusion, and market equilibrium (Cachon et al., 2017; Yildiz & Savelsbergh, 2019; Benjaafar & Hu, 2020; Bernstein et al., 2021; Fatehi & Wagner, 2022; Miao et al., 2022; Tripathy et al., 2022). Predicted demand and delivery capacity (supply) correspond to variables unavailable for confidentiality

reasons: Alpha could not disclose the predicted demand for delivery service requests, nor the current (actual) capacity of drivers. Although we could observe the actual demand (i.e., how many delivery task were performed), we decided not to use it as instrumental variable because potential source of endogeneity for reverse causality. Indeed, delivery task remuneration would directly impact the number of completed tasks on a given day. As discussed in the variable operationalization section, we assume that the driver who accepted and performed the delivery task was active (i.e., had its phone available) at the time the offer was broadcasted, thus observing only the actual delivery capacity. Finally, to verify that this source of endogeneity does not affect the estimations, we consulted Alpha operations senior managers. On this matter, they confirmed that the platform strives to "*…maintain the zone-level earnings per hour…*" to avoid "*…a bump in offer payout….*" However, the potential presence of *bumps* due to variations in demand and supply are not captured in the dataset and we acknowledge as a source of endogeneity.

Second, the extension of the service area influence the remuneration for a delivery task (Castillo et al., 2022a). Platforms typically establish a desired trade-off for cost/service levels for service areas (Shen & Daskin, 2005; Castillo et al., 2022a), which depends on the number of depots in the area (Fatehi & Wagner, 2022), the type of delivery area (i.e., urban vs rural) (Janjevic & Winkenbach, 2020), and the predefined set of available delivery time (Amorim et al., 2020). Although the dataset reports the unique id for driver delivery zones, we could not capture the desired level of service for each delivery zone/area.

Third, the compensation scheme used by the crowdshipping platform is another element influencing delivery task remuneration (Arslan et al., 2019; Dai & Liu, 2020; Dayarian & Savelsbergh, 2020). We exclude that variance in the compensation scheme constitutes a source of endogeneity because the observations of the delivery tasks remuneration were collected from a

single platform that utilizes the same scheme across all markets. Aligned with other crowdshipping platforms operating in the same space (i.e., last mile delivery) (Arslan et al., 2019), Alpha utilizes a hourly rate as compensation scheme. Hence, aspects such as the number of orders within the delivery task, the clock time of the day and day of the week (traffic) (Li et al., 2022), and the characteristics of the delivery zone may influence the hourly rate and the remuneration for a delivery task.

Therefore, to overcome the challenges of endogeneity for omitted variable bias, we adopted a IV/2SLS econometric approach (Ho et al., 2017; Lu et al., 2018). This ensures to identify the causal estimation for the effect of the focal predictors on the dependent variables. The first step in the IV/2SLS approach is to identify instrumental variables that meets the relevance and exclusion conditions, per Wooldridge (2010). Indeed, an instrument must explain the suspected endogenous predictor (relevance condition), and affect the outcome variables only through the focal predictor being unrelated to the unobservable variables included in the error term (Wooldridge, 2016). Hence, in the context of this study, such instrument must relate to delivery task remuneration and Tip%, without affecting driver behaviors relative to a specific delivery task.

We identify the instrument for delivery task remuneration as the average delivery task remuneration per zip code p, day of the week w, clock hour time of the day h, and number of orders in the delivery task o ($avgM_{pwho}$). In other words, to address the potential source of endogeneity related to the unobserved desired level of cost/service for a delivery area and the characteristics that the compensation scheme utilizes to compute the delivery task remuneration for each delivery task remuneration, we computed the mean of the subgroup of which the delivery task pertains. Similar operationalizations of instrumental variables based on the average of identified subgroups within the dataset has been used in the operations management literature

(Cachon, Gallino, & Olivares, 2019; Dhanorkar & Siemsen, 2021; Akturk, Mallipeddi, & Jia, 2022). It is conceivable that the average of delivery task remunerations for subgroups of delivery tasks represents a strong instrument because it highly correlates to the suspected endogenous predictor, and exogenous instrument given that delivery drivers cannot reasonably know or predict the average compensation for all delivery tasks belonging to a subgroup. Following Akturk et al. (2022)'s reasoning, it is unlikely that an individual observes the actual platform's prior, current, and future remunerations for each delivery task performed within a subgroup, and could only infer an average based on personal experience.

Estimation Models

Depending on the outcome variable of interest and the unit of analysis, we tested the hypotheses using the appropriate IV/2SLS estimation model. Hence, for the outcome variables A_i, ST_i, we used a 2SLS approach with the unit of analysis as delivery task i. We estimated the predicted values of the two endogenous predictors in the first-stage of the 2SLS. We ran an OLS regression with delivery task remuneration as the outcome variable, and the control variables and instrumental variables as predictors:

(3) $M_i = \beta_0 + \beta_1 avgM_{pwho} + \mathrm{BX}_i + \varepsilon_{id}$

Where X_i is the vector of control variables, B is the vector of coefficients for the control variables, and ε_{id} are robust standard errors, clustered for driver id to mitigate the potential source of heteroscedasticity. BX_i is specified as follows:

(4) $\mathrm{X}_i = \beta_0 + \beta_1 cum_d + \beta_2 exp_d + \beta_3 dist_i + \beta_4 age_d + \beta_5 ndr_z + \beta_6 nor_i + \beta_7 nun_i +$
$\beta_8 pci_p + \beta_{9is} store_{is} + \beta_{10im} month_{im} + \beta_{11iw} dow_{iw} + \beta_{12it} time_{it} + \beta_{13}\mathrm{imr}$

We indicate the corresponding predicted values as $\widehat{M_i}$ and $\widehat{T_i}$, respectively, that we replace to M_i and T_i when estimating model for the outcome variables. For the given outcome variables $Y_i =$ (ln $(A_i), ST_i)$, the final estimation models were specified as follows:

(5) $Y_i = \alpha_0 + \alpha_1 \widehat{M_i} + \mathrm{BX}_i + \varepsilon_{id}$

To test the hypotheses for the outcome variable $DRET_d$ (the driver return time), we aggregated the cleaned dataset into a panel dataset with delivery drivers' ids as the cross-sectional unit, and the calendar day as unique time id, resulting in ~800k observations. The panel dataset reports the sum of delivery task remuneration that a single driver earned on a given day. Similarly to the other outcome predictors, we also identified the instrumental variables for the two predictors by computing the average of delivery task remuneration $(avgM_{pwo})$ for a given zip code p, day of the week w, and number of orders o. Then, we estimated the first and second stage of the IV/2SLS approach following (Wooldridge, 2010). In the first stage, we regressed the daily delivery task remuneration on the two instrumental variables as follows:

(6) $M_{dy} = \beta_0 + \beta_1 avgM_{pwo} + \mathrm{BX}_{dy} + \varepsilon_{dy}$

Where the subscript y indicates the calendar day, the ε_{dy} indicates the robust standard errors, clustering for driver id to mitigate potential source of heteroscedasticity, and BX_{dy} was specified as follows:

(7) $\mathrm{X_{dy}} = \beta_0 + \beta_1 exp_{dy} + \beta_2 ntask_{dy} + \beta_3 ndr_y + \beta_{4ym} month_{ym} + \beta_{5yw} dow_{yw} + \beta_{6dyt} time_{dyt}$

Where exp_{dy} indicates the driver experience accumulated since the beginning of the dataset, $ntask_{dy}$ indicates the total number of delivery tasks completed by driver d on the day y, ndr_y indicates the number of active drivers on a given day d, and $month_{ym}$, dow_{yw}, and $time_{dyt}$ represent the time fixed effects for month, day of the week, and the percentage of delivery task

completed during the morning, afternoon, and afterhours in the day in which the driver performed the delivery task. We indicate the corresponding predicted values as $\widehat{M_{dy}}$, that we replace to M_{dy} when estimating the model for the outcome variables. The final estimation models for $DRET_d$ was specified as follows:

$$(8) \quad DRET_d = \alpha_0 + \alpha_1 \widehat{M_{dy}} + \mathrm{BX}_{dy} + \varepsilon_{dy}$$

Challenges of working with large samples

Another potential challenge that we face in this study is the large sample size. While large datasets helps generating more precise estimates (Ho et al., 2017), we discuss two challenges related to establishing statistical inference when using large samples. Large samples include more than 10,000 observations per Lin, Lucas, & Shmueli (2013).

The first challenge is that p-value quickly approaches zero as the sample size increases. Hence, researchers can draw incorrect statistical inferences based on a significant p-value that is driven by the large sample (Mohajeri, Mesgari, & Lee, 2020). Lin et al. (2013) suggest several remedies, including reporting the effect size and the confidence interval for each parameter. The effect size reflects the sensitivity of the dependent variable to changes in the independent variable. Reporting the most conservative bound of confidence intervals at 95% inform about the range and magnitude of the parameters, especially with large sample sizes in which confidence intervals tend to be tighter. Hence, following prior literature, for each estimation model we reported effect sizes and confidence intervals for all parameters.

The second challenge is the setting of p-value thresholds. Benjamin et al. (2018) discussed that the commonly accepted threshold of $p<.05$ results in a high rate of false positives and recommended adopting a less conservative threshold of $p<.005$. Hence, following Benjamin et al. (2018), we cautiously interpreted the estimations by looking at the level of significance.

Results and Discussion

We first presented statistical evidence relative to our claims for endogeneity. Then, we reported the testing of each hypothesis and a plot for each significant interaction effect. For linear estimations, the plots use low and high values of continuous variables computed as $\pm$ one standard deviation from the mean. We also included additional analyses about VIF (Perdikaki et al., 2015), effect size, and confidence intervals for each estimated coefficient (Lin et al., 2013; Benjamin et al., 2018). Furthermore, similar to prior literature (Abbey et al., 2015), we reported the standardized beta coefficient to ensure a more reliable comparison between effects on the same outcome variable.

Tests for endogeneity

We assessed the presence of endogeneity and performed the appropriate statistical tests to ensure that the instruments were strong. Specifically, for the dependent variables A_i, ST_i, we report the Cragg-Donald Wald F-test for the first stage regression of each pair of instrumental variables and endogenous predictors. Then, we report the Durbin χ^2 for the Durbin-Wu-Hausman test that compares the OLS estimations with the instrumental variable ones (Lu et al., 2018). Table 4 reports the statistical evidence that (1) the instruments are strong and (2) the suspected endogenous focal predictor is, indeed, endogenous.

Table 4 – Test for endogeneity

Suspected endogenous variable	Delivery Task Remuneration	
Instrument	$avgM_{pwho}$	
DV	Durbin χ^2	Cragg-Donald Wald F-test
ln (A_i)	497.35 $p < .01$	704,440.10 $R^2 = .30$ $p < .01$
ST_i	179.63 $p < .01$	704,440.10 $R^2 = .30$ $p < .01$

Next, we reported the results of the first stage in Table 5. Model 1 reports the estimations of equation 3 for the impact of the instrumental variable $avgM_{pwho}$ on delivery task remuneration (β = 0.989, SE = 0.001).

Table 5 – IV/2SLS results of the first stage

	(1) M_i β	se
Cum n del	-0.108****	(0.005)
Driv exp	-0.003****	(0.001)
Miles prev stop to new pickup	0.018	(0.045)
Driv age	-0.012****	(0.001)
N driv in zone	0.005****	(0.001)
N orders in batch	-0.924****	(0.015)
Del unattended	2.541****	(0.018)
Inverse Mills Ratio	0.036	(0.045)
Store fixed effects	YES	
Month fixed effects	YES	
Day of the week fixed effects	YES	
Time of the day fixed effects	YES	
$avgM_{pwho}$	0.989****	(0.001)
$avgT_{pwho}$		
Constant	-0.566****	(0.068)
r2	0.299	
r2 adjusted	0.299	
N	1,667,993	
F	35340.185****	

Note: $^{+}$ $p < 0.1$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.005$, **** $p < 0.001$. Reported robust standard errors are clustered on driver ID.

For the dependent variable $DRET_d$, we conducted endogeneity tests to assess whether the suspected endogenous focal predictors were indeed endogenous (Ho et al., 2017; Lu et al., 2018). First, we ran a fixed-effect regression with xtivreg2 on STATA 17. The first-stage results supported the notion that the instruments meet the exclusion condition. Specifically, the Sanderson-Windmeijer chi-squared and F-test, Anderson-Rubin Wald test, and the Stock-Wright LM S-statistic reject the null hypothesis that the weakly identified (Lu et al., 2018). Second, the Davidson-McKinnon test retained the null hypothesis of an exogenous predictor (Wooldridge, 2010; Chuang, Oliva, & Perdikaki, 2016; Liu et al., 2016), suggesting that endogeneity is not a concern in this analysis and an OLS regression can be adopted. Finally, we also compared fixed-

effect to random-effect estimations through a Hausman test (Hausman, 1978; Certo, Withers, & Semadeni, 2017), which reveals that a fixed-effect estimation is the most appropriate. Table 6 reports the statistical tests addressing endogeneity.

Table 6 – Test for endogeneity

Suspected endogenous variable	Daily delivery Task Remuneration
Instrument	$avgM_{pwo}$
Sanderson-Windmeijer χ^2	44,418.70 $p < 0.01$
Sanderson-Windmeijer F-test	44,417.91 $p < 0.01$
Anderson-Rubin Wald test	3.73 $p < 0.05$
Stock-Wright LM S-statistic	3.73 $p < 0.05$
Davidson-McKinnon test	0.38 $p = 0.54$
Hausman χ^2 comparing fe and re	3,330.79 $p < 0.01$

Effects on Pre-task Behavior - Acceptance response time

To test hypotheses H1a, H4a, and H5a, we report in Table 7 the IV/2SLS second stage estimations relative to the outcome variable Acceptance response. Model 1 of Table 7 tested the following estimation model:

(9) $\ln(A_i) = \alpha_0 + \alpha_1 \widehat{M_i} + \mathrm{BX}_i + \varepsilon_{id}$

H1 predicted a positive effect of remuneration on driver pre-task behaviors, with a decrease in Acceptance response time at higher levels of remuneration. Delivery task remuneration presents a negative significant coefficient (α = -0.027, SE = 0.001) on Acceptance response time. Hence, we find support for H1.

Next, Model 3 reports the coefficients for the interaction effects by testing this model:

(10) $\ln(A_i) = \alpha_0 + \alpha_1 \widehat{M_i} + \alpha_2 h_i + \alpha_3 E_i + \alpha_4 \widehat{M_i} \times h_i + \alpha_4 \widehat{M_i} \times E_i + \mathrm{BX}_i + \varepsilon_{id}$

H4a suggested that drop-off density enhances the negative relationship between remuneration and acceptance response time. The interaction effect shows a negative significant interaction coefficient (α = -0.005, SE = 0.001). We plot the estimated coefficients in Figure 8, which shows that density decreases acceptance response time, respectively, by 18% at high levels

of remuneration, and 6% at low levels of remuneration, supporting H4a. Comparing the highest (low remuneration – low density) and lowest (high remuneration – high density) levels in the plot by exponentiating the relative values (1.98 and 1.09), we notice that Acceptance response time decreases by 5 minutes.

H5a suggested that an expedited delivery mitigates the relationship between remuneration and pre-task behaviors. Model 3 reports a positive significant interaction coefficient (α = 0.009, SE = 0.001), and Figure 9 reveals that an expedited delivery is less preferable to a standard one with an increase of 26% on acceptance response time at high levels of Remuneration, but only an increase of 7% at low levels of remuneration. Further, we notice that the slope for standard delivery is steeper than for expedited ones, suggesting that expedited delivery mitigates the negative effect of remuneration, and supports H5a.

Table 7 – IV/2SLS second stage results for Acceptance response time and delivery task remuneration

	(1) ln (A_i) β	se	(2) ln (A_i) β	se	(3) ln (A_i) β	se
cum_{fd}	-0.082****	(0.002)	-0.083****	(0.002)	-0.083****	(0.002)
exp_d	0.000	(0.000)	0.001**	(0.000)	0.001**	(0.000)
$dist_i$	0.001	(0.009)	0.001	(0.009)	0.001	(0.009)
age_d	-0.005****	(0.000)	-0.005****	(0.000)	-0.005****	(0.000)
ndr_z	-0.005****	(0.001)	-0.003****	(0.001)	-0.003****	(0.001)
nor_i	0.235****	(0.005)	0.395****	(0.006)	0.395****	(0.006)
nun_i	0.053****	(0.003)	-0.001	(0.003)	-0.001	(0.003)
imr	-1.401****	(0.021)	-1.507****	(0.021)	-1.509****	(0.021)
Store fe	YES		YES		YES	
Month fe	YES		YES		YES	
Day of the week fe	YES		YES		YES	
Time of the day fe	YES		YES		YES	
Remuneration	-0.027****	(0.000)	-0.027****	(0.000)	-0.028****	(0.000)
Density			-0.200****	(0.003)	-0.208****	(0.003)
Expedited			0.282****	(0.007)	0.314****	(0.008)
Remuneration x Density					-0.005****	(0.000)
Remuneration x Expedited					0.009****	(0.001)
Constant	1.684****	(0.030)	1.566****	(0.029)	1.565****	(0.029)
r2	0.081		0.086		0.086	
r2 adjusted	0.081		0.086		0.086	
N	1,667,993		1,667,993		1,667,993	
F	1679.703****		1638.167****		1512.995****	

Note: A_i is Acceptance response time (minutes). $^{+}$ $p < 0.1$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.005$, **** $p < 0.001$. Reported robust standard errors are clustered on driver ID.

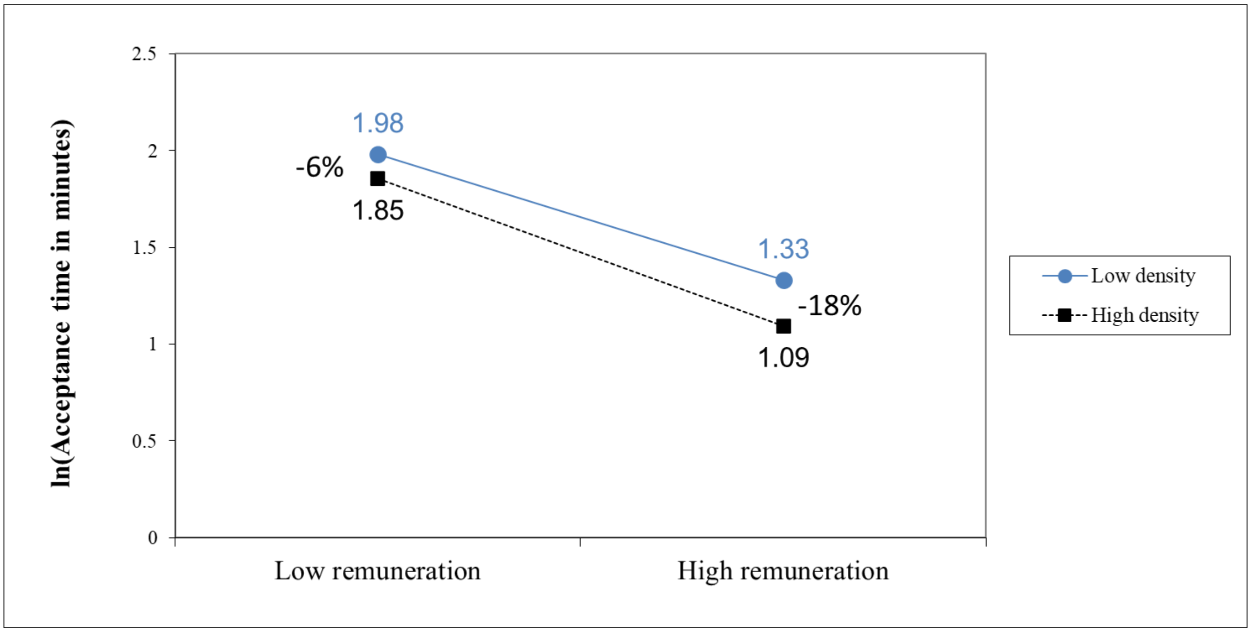


Figure 8 – Interaction between Remuneration and Density on Acceptance response time

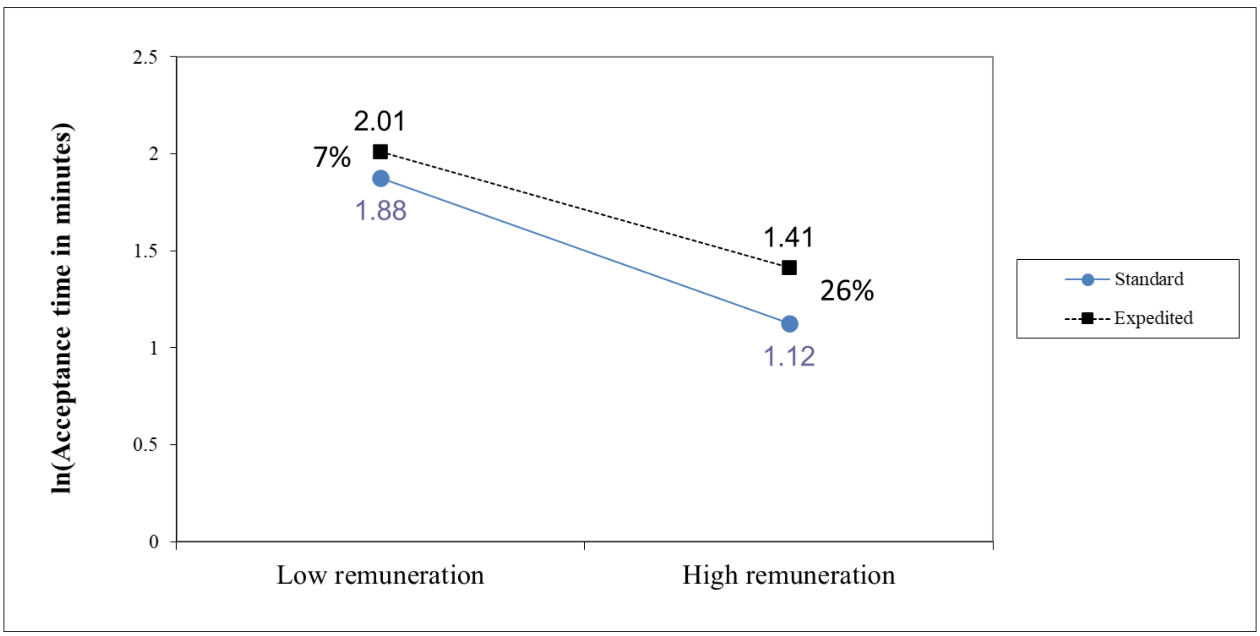


Figure 9 – Interaction between Remuneration and Expedited on Acceptance response time

As recommended in prior literature, Table 8 reports the beta standardized coefficients, the vif, the confidence intervals at 95%, and the effect size η^2 for each of the coefficients reported in Table 7 for model 1 (testing H1a) and model 3 (testing H4a and H5a).

Table 8 – Complementary statistics to IV/2SLS second stage results for Acceptance response time and delivery task remuneration

	(1) ln (A_i) Beta	VIF	CI 95%	η²	(3) ln (A_i) Beta	VIF	CI 95%	η²
cum_{fd}	-0.069	(1.060)	[-0.086,-0.079]	0.006	-0.07	(1.061)	[-0.087,-0.080]	0.006
exp_d	0.001	(1.807)	[-0.000,0.001]	0.001	0.008	(1.815)	[0.000,0.001]	0.001
$dist_i$	-0.029	(1.000)	[-0.016,0.017]	0.001	-0.029	(1.000)	[-0.016,0.018]	0.001
age_d	-0.026	(1.015)	[-0.006,-0.004]	0.001	-0.018	(1.015)	[-0.006,-0.004]	0.001
ndr_z	0.062	(1.141)	[-0.006,-0.004]	0.001	0.104	(1.155)	[-0.004,-0.002]	0.001
nor_i	0.015	(1.231)	[0.225,0.244]	0.004	-0.029	(1.686)	[0.384,0.407]	0.005
nun_i	-0.189	(1.237)	[0.046,0.060]	0.001	-0.028	(1.275)	[-0.008,0.005]	0.001
imr	-0.09	(1.355)	[-1.442,-1.361]	0.029	-0.204	(1.406)	[-1.550,-1.467]	0.031
Remuneration	-0.029	(1.159)	[-0.028,-0.026]	0.008	-0.094	(1.310)	[-0.029,-0.027]	0.008
Density					-0.077	(1.257)	[-0.215,-0.202]	
Expedited					0.048	(1.897)	[0.298,0.330]	0.004
Remuneration x Density					-0.01	(1.125)	[-0.006,-0.004]	0.001
Remuneration x Expedited					0.008	(1.632)	[0.007,0.011]	0.001

Note: Beta coefficient, VIF, CI 95%, and η² (Effect Size) are based on the unstandardized coefficients reported in Table 7.

We performed a series of sensititivty analyses to ensure the robustness of our estimations (Appendix A).

Discussion of results for the effects on pre-task behaviors – Acceptance response time

This set of results suggests that the likelihood of accepting a delivery task depends not only on the monetary incentives, but also on the operational characteristics of the delivery task. Following SET and IOE, before accepting an exchange, individuals identify the reward from such exchange, which in the just-in-time scheduling is also based on the effort a server would need to commit to complete the exchange. Denser delivery tasks represent a better income opportunity and reward from the exchange for a delivery driver, who would guarantee him/herself a greater remuneration for a lower commitment in terms of effort and time. Conversely, an expedited delivery requires drivers to accept and complete the task with short notice, change their delivery plans, and renounce the flexibility of the gig, generating a work-life conflict and a perceived lack of control (Kamalahmadi et al., 2021). Delivery drivers will be reluctant to accept an expedited delivery, even at high levels of remuneration or pre-tipping.

These results offer several contributions. First, we contribute to the growing literature investigating the central issue of crowdshipping of supply uncertainty (Benjaafar & Hu, 2020; Dayarian & Savelsbergh, 2020). Adding to this literature that primarily investigated the effect of monetary incentives on scheduling capacity (Cachon et al., 2017; Castillo et al., 2018, 2022b; Duhaime & Woessner, 2019; Bernstein et al., 2021; Miao et al., 2022), this study unveils the fundamental role of operational characteristics of the delivery task impacting supply capacity. Hence, we acknowledge that monetary incentives serve as the main predictor to engage crowdshipping delivery drivers (e.g., greater remuneration decreases Acceptance response time), but we contribute by identifying the role of task characteristics. Relatedly, we also contribute to the literature investigating last mile delivery fleet design, which finds an optimal solution in a hybrid system that balances private and crowdshipping fleets to ensure both on-demand service and consistency in the service level (Dai & Liu, 2020; Castillo et al., 2022a). This study adds the importance of operational characteristics of the task to improve the stability of the supply capacity and the preferred level of service to the customer.

Second, we contribute to SET and IOE by investigating the several aspects of an exchange and income opportunity that delivery drivers evaluate before committing to such exchange. Following the contribution to the crowdshipping literature, we inform SET relative to the several aspects of an exchange that individuals (delivery drivers) evaluate. Indeed, the results of this study extend SET success proposition (Homans, 1958; Narasimhan et al., 2009) by focusing on not only the reward but also on the perceived level of effort that the exchange will require. Hence, the mere reward (i.e., remuneration) is necessary but not sufficient to ensure the promised commitment to the exchange, because delivery drivers will also account for the time and effort for the task *before* accepting the exchange. A related aspect is the exchange partner's strategy to reduce such

perceived effort. At the same level of remuneration, drivers will more favorably engage in exchanges for which the effort was somehow reduced by the platform (greater density by consolidating drop-offs into the same delivery area) or by the customer (standard delivery vs expedited delivery).

Next, this study contributes to SET by focusing on the exchange partners. Before acceptance, a delivery driver identifies the exchange rewards with two separate parties: the platform, which provides the compensation, and the customer, who promises a tip. IOE suggests that larger gratuities constitute an additional income opportunity (Kamalahmadi et al., 2021), and (Castillo et al., 2022b) confirms that tipping mitigates uncertainty in the service level in the crowshipping context. Our contribution to SET is the evolution of the initial platform-driver exchange into platform-driver-customer exchange, with both the platform *and* the customer playing an active role in decreasing the service provider uncertainty. Therefore, we extend SET value proposition (Emerson, 1962; Narasimhan et al., 2009) by suggesting that an exchange may results in multiple exchanges, each resulting in distinct counterpart's actions. For delivery drivers, the value of the exchange with the platform is not as important as the value from the exchange with the customer.

Effects on Task behaviors - Service time

To test hypotheses H2a, H4b, and H5b, relative to the impact of delivery task remuneration, the moderators density, expedited, and interaction terms on task behavior of Service time, we report in Table 9 the IV/2SLS second stage estimations. Model 1 in Table 9 tests the following estimation model:

$$(11)\quad ST_i = \alpha_0 + \alpha_1 \widehat{M_i} + \mathrm{B}\mathrm{X}_i + \varepsilon_{id}$$

H2a predicted a positive effect of remuneration on Driver task behaviors, with a decrease of Service time at higher levels of remuneration. Remuneration presents a positive significant coefficient (α = 0.072, SE = 0.002) on Service time. Thus, H2a was not supported.

Next, Model 3 reports the coefficients for the interaction effects by testing this model:

(12) $ST_i = \alpha_0 + \alpha_1\widehat{M_i} + \alpha_2 h_i + \alpha_3 E_i + \alpha_4\widehat{M_i} \times h_i + \alpha_5\widehat{M_i} \times E_i + \mathrm{BX}_i + \varepsilon_{id}$

H4b suggested that higher drop-off density enhances the positive relationship between remuneration and Driver task behaviors, hence enhancing the negative effect. The interaction effect shows a negative significant interaction coefficient (α = -0.051, SE = 0.002). Figure 10 reveals two important results. First, delivery density significantly reduces the Service time needed to complete the task independently of task remuneration, with a decrease between 21% and 36% of Service time depending on the remuneration. The importance of this result is a difference of approximately 10 minutes (-36%) between delivery tasks having the same level of high remuneration, but differentiated by the density of the task. In the crowdshipping context with on-demand deliveries from stores to the local neighborhood, 10 minutes is often the Service time (we recall that Service time is the time it takes to a driver to complete the task since s/he leaves the store) of an additional task that could have been performed. Second, density moderates the relationship between remuneration and Service time in different ways: Higher levels of density reduce Service time with greater remuneration, whereas lower levels of density increase Service time with greater remuneration. Thus, H4a was partially supported.

H5b suggested that an expedited delivery enhances the effect of remuneration on Service time. Model 3 reports a non-significant coefficient, thus, we do not find support H5b.

Table 9 – IV/2SLS second stage results for Service time and delivery task remuneration

	(1)		(2)		(3)	
	ST_i		ST_i		ST_i	
	β	se	β	se	β	se
cum_{fd}	-0.276****	(0.005)	-0.251****	(0.005)	-0.252****	(0.005)
exp_d	-0.018****	(0.001)	-0.015****	(0.001)	-0.015****	(0.001)
$dist_i$	0.161****	(0.026)	0.125****	(0.024)	0.124****	(0.023)
age_d	0.031****	(0.002)	0.028****	(0.002)	0.027****	(0.002)
ndr_z	-0.010****	(0.001)	0.024****	(0.001)	0.026****	(0.001)
nor_i	10.319****	(0.037)	10.490****	(0.041)	10.594****	(0.042)
nun_i	-1.763****	(0.018)	-1.752****	(0.016)	-1.755****	(0.016)
imr	-1.805****	(0.087)	-2.051****	(0.082)	-2.082****	(0.081)
Store fe	YES		YES		YES	
Month fe	YES		YES		YES	
Day of the week fe	YES		YES		YES	
Time of the day fe	YES		YES		YES	
Remuneration	0.072****	(0.001)	0.048****	(0.001)	0.043****	(0.001)
Density			-3.709****	(0.012)	-3.794****	(0.013)
Expedited			-1.189****	(0.022)	-1.150****	(0.025)
Remuneration x Density					-0.051****	(0.002)
Remuneration x Expedited					0.005+	(0.003)
Constant	6.279****	(0.095)	6.707****	(0.088)	6.725****	(0.088)
r2	0.203		0.305		0.306	
r2 adjusted	0.203		0.305		0.306	
N	1,667,993		1,667,993		1,667,993	
F	8575.988****		13163.886****		12387.741****	

Note: ST_i is Service time (minutes). + $p < 0.1$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.005$, **** $p < 0.001$. Reported robust standard errors are clustered on driver ID.

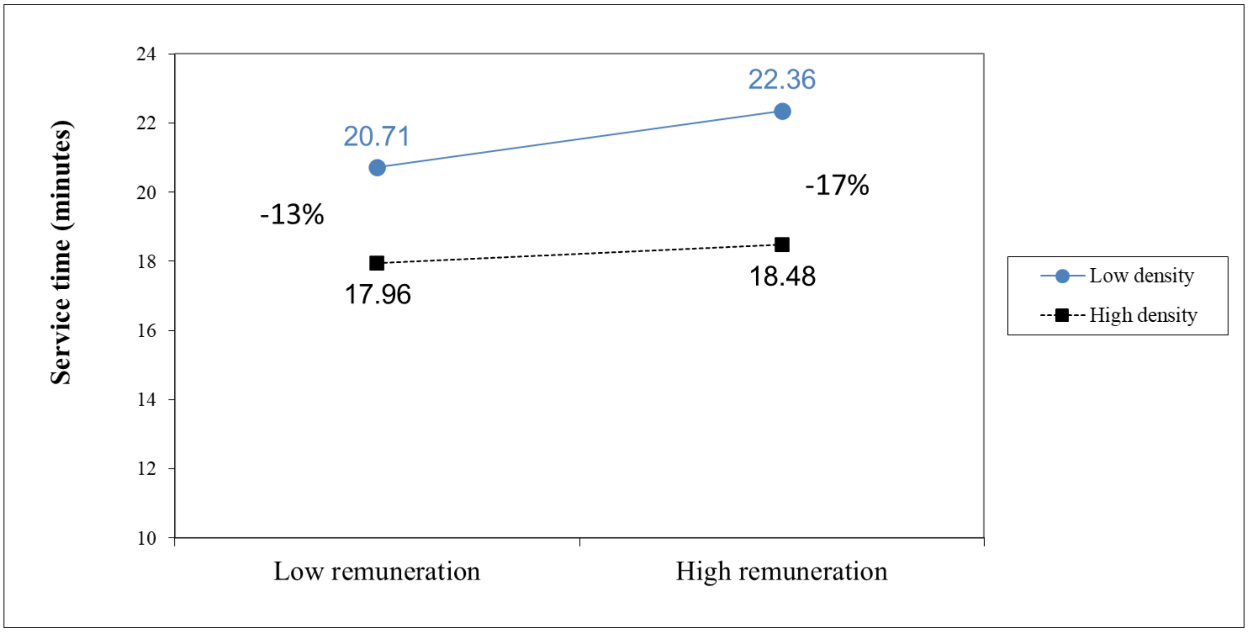


Figure 10 – Interaction between Remuneration and Density on Service time

As for Acceptance response time, Table 10 report the beta standardized coefficients, the vif, the confidence intervals at 95%, and the effect size η^2 for each of the coefficients reported in Table 9.

Table 10 – Complementary statistics to IV/2sls second stage results for Service time and delivery task remuneration

	(1) ST_i Beta	VIF	CI 95%	η^2	(3) ST_i Beta	VIF	CI 95%	η^2
cum_{fd}	-0.052	(1.060)	[-0.244,-0.222]	0.006	-0.050	(1.061)	[-0.239,-0.218]	0.465
exp_d	-0.046	(1.807)	[-0.016,-0.013]	0.002	-0.032	(1.815)	[-0.011,-0.008]	0.007
$dist_i$	0.002	(1.000)	[0.038,0.170]	0.001	0.002	(1.000)	[0.029,0.151]	0.002
age_d	0.029	(1.015)	[0.019,0.025]	0.002	0.027	(1.015)	[0.017,0.023]	0.001
ndr_z	-0.018	(1.141)	[-0.013,-0.007]	0.001	0.023	(1.155)	[0.018,0.023]	0.002
nor_i	0.05	(1.231)	[0.498,0.564]	0.17	0.119	(1.686)	[1.504,1.588]	0.002
nun_i	0.393	(1.237)	[5.697,5.764]	0.009	0.313	(1.275)	[4.781,4.844]	0.191
imr	-0.037	(1.355)	[-0.999,-0.719]	0.003	-0.058	(1.407)	[-1.694,-1.425]	0.012
Remuneration	0.023	(1.159)	[0.070,0.075]	0.006	0.039	(1.310)	[0.041,0.046]	0.185
Density					-0.348	(1.257)	[-3.819,-3.770]	0.002
Expedited					-0.041	(1.897)	[-1.199,-1.101]	0.008
Remuneration x Density					-0.025	(1.125)	[-0.054,-0.048]	0.007
Remuneration x Expedited					-0.027	(1.632)	[-0.001,0.011]	0.001

Note: Beta coefficient, VIF, CI 95%, and η^2 (Effect Size) are based on the unstandardized coefficients reported in Table 9.

We performed a series of sensititivity analyses to ensure the robustness of the estimations (see Appendix A).

Discussion of results for the effects on task behaviors – Service time

This set of results provide support that in contrast to our predictions, delivery task remuneration does not incentivize delivery driver task behaviors relative to Service time. We explain this result following two distinct, yet complementary explanations. First, aligned with the logic related to the compensation scheme (Arslan et al., 2019), remuneration presents a positive effect on Service time because delivery drivers may be paid more to perform longer. Second, drivers would take longer to perform a task because, knowing that they will be receiving such high compensation, they would not put an extra effort to perform a task faster, switching the value of the exchange to their advantage. In addition, delivery task operational characteristics would mitigate such effect

reducing the Service time, especially for high density and expedited, yet drivers completing a complex task receiving a high remuneration would take longer to perform the task because of the compensation scheme as well as no extra effort required to task completion.

Second, we contribute to SET and IOE. Social exchange theory typically do not predict an exchange part performance. However, IOE extends SET by offering the theoretical background to investigate the impact of incentives and rewards on a server performance in the just-in-time scheduling. This study contributes to SET by suggesting that depending on the type of reward, and which exchanging parts provide the reward, the performance will likely improve. Hence, it is important to clarify the role of each exchange part: despite the platform offers the reward and strives to ensure the service quality, the reward for the exchange with the customer is what matters the most. This study also contributes to IOE. While IOE studies the effect of the server's intrinsic abilities moderating the impact of monetary incentives (income opportunity effect) on servers' performance, we expand IOE by studying the intrinsic characteristics of the delivery task. Hence, in a context of just-in-time scheduling with independent contractors as servers, a platform that cannot control which driver will be accepting the task (in this context, the platform just broadcasts the offer onto the Driver App), attempts to improve the performance through the intrinsic characteristics of the task.

Effects on Post-task behaviors – Driver Satisfaction and Retention

The final set of hypotheses considers the impact of delivery task remuneration and the interaction effects with delivery task operational characteristics on post-task behaviors of driver retention. To study the hypotheses H3, H4c, and H5c relative to the impact of delivery task remuneration, the moderators density, expedited, and interaction terms on driver retention in the form of driver return time, we report in Table 11 the OLS regression estimations obtained from the panel dataset. We

recall that we found statistical evidence to support the notion that the total delivery remuneration for a given day of deliveries does not present endogeneity concerns on the number of days until the driver performs another delivery task.

Model 1 in Table 11 tests the following estimation fixed-effect model:

(13) $DRET_{dy} = \alpha_0 + \alpha_1 M_{dy} + \mathrm{BX}_{dy} + \varepsilon_{dy}$

H3a predicted a positive effect of remuneration on driver post-task behaviors, with a decrease of number of days at higher levels of remuneration. Remuneration day presents a positive and significant coefficient (α = 0.001, SE = 0.001) on driver return time. Thus, H3a was not supported.

Next, Model 3 reports the coefficients for the interaction effects by testing this model:

(14) $DRET_{dy} = \alpha_0 + \alpha_1 M_{dy} + \alpha_2 h_{dy} + \alpha_3 E_{dy} + \alpha_4 M_{dy} \times h_{dy} + \alpha_5 M_{dy} \times E_{dy} + \mathrm{BX}_{dy} + \varepsilon_{dy}$

Where, differently from the previous analyses, h_{dy} refers to the density of all delivery tasks performed in a given day (total number of orders over total number of miles), E_{dy} represents the number of expedited delivery tasks completed in a given day.

H4c suggested that higher drop-off density enhances the positive relationship between remuneration and retention. The interaction effect shows a non-significant coefficient. Hence, H4c was not supported. H5c suggested that a larger number of expedited deliveries mitigates the effect of remuneration on retention. Model 3 reports a positive and significant interaction term (α = 0.001, SE = 0.001), and Figure 11 shows that at high levels of remuneration, the difference in return days is smaller between standard and expedited as compared to lower levels of remuneration, thus supporting H5c.

Table 11 – OLS estimations for Driver return time and daily remuneration

	(1) $DRET_{dy}$ β	se	(2) $DRET_{dy}$ β	se	(3) $DRET_{dy}$ β	se
exp_{dy}	-0.019****	(0.000)	-0.019****	(0.000)	-0.019****	(0.000)
$ntask_{dy}$	-0.078****	(0.004)	-0.077****	(0.004)	-0.078****	(0.004)
ndr_{y}	0.001****	(0.000)	0.001****	(0.000)	0.001****	(0.000)
Month fe	YES		YES		YES	
Day of the week fe	YES		YES		YES	
Time of the day fe	YES		YES		YES	
Remuneration day	0.001****	(0.000)	0.001****	(0.000)	0.001****	(0.000)
Density			0.048***	(0.016)	0.043*	(0.021)
Expedited			-0.000	(0.004)	-0.026****	(0.008)
Remuneration day x Density					0.000	(0.000)
Remuneration day x Expedited					0.000****	(0.000)
Constant	3.409****	(0.017)	3.390****	(0.018)	3.418****	(0.019)
Within r2	0.015		0.015		0.015	
N	862,199		862,199		862,199	
F	501.593****		413.375****		352.713****	

Note: $DRET_{dy}$ is driver return time (days). $^{+}$ $p < 0.1$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.005$, **** $p < 0.001$. Reported robust standard errors are clustered on driver ID.

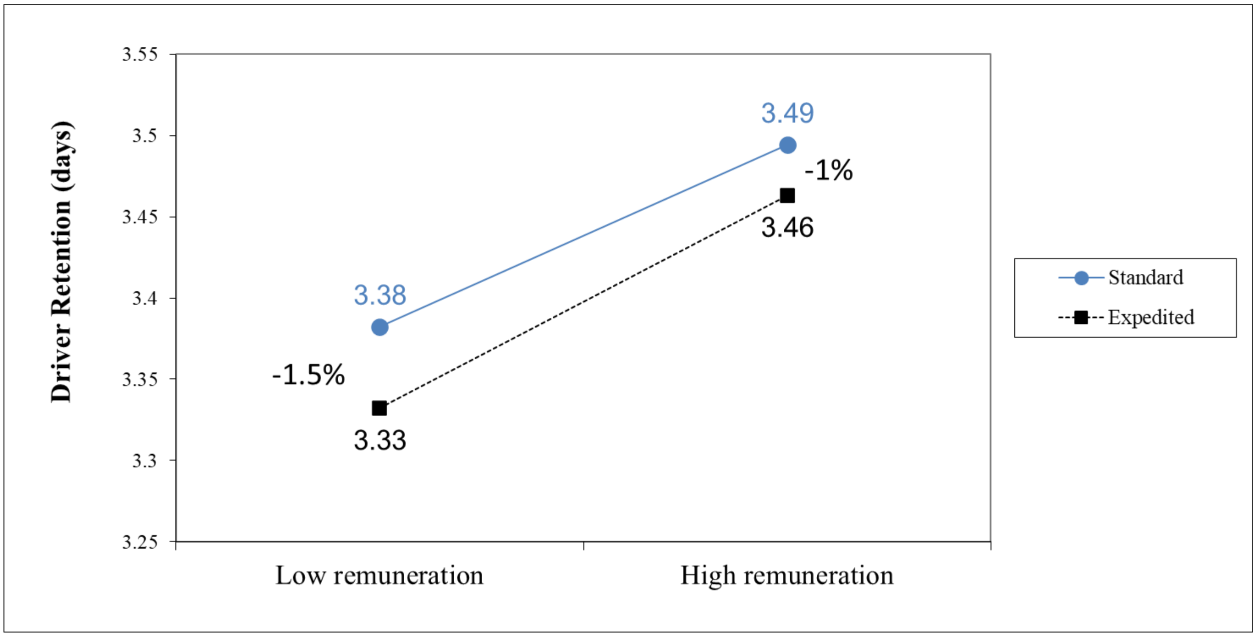


Figure 11 – Interaction between Remuneration and Expedited on Retention

Survival analysis

An additional analysis that is typically carried out in the literature to assess driver retention focuses on investigating the effect of the focal predictors on Driver attrition (Castillo et al., 2022b). This is of particular interest in the crowshipping context, where there is a debate relative to the real

consequences of driver attrition. This context is characterized by high attrition rates, with evidence of 60% of drivers no longer active after six months (Rosenblat, 2016; Kooti et al., 2019; Cook et al., 2021). Some support the notion that the number of newcomers outweighs the attrition rate, leading to a growth of the platform (Kooti et al., 2019). In contrast, others sustain that such high attrition rates affect the feasibility of desired service levels and platform long-term strategy (Rosenblat, 2016). This is explained by the literature discussing network effects in crowdshipping. The network effect occurs in the early stages of market penetration for a given platform, and manifests in the iterative loop between the number of drivers and the number of customers (Apte & Davis, 2019). Greater is the number of drivers, cheaper is the service and easier the matching, attracting more customers, and in turn more drivers (Mittal et al., 2021). According to existing literature, such network effects tend to exist only at local level (Cullen & Farronato, 2021), it is not scalable in the short term with surges in demand for delivery service (Qi et al., 2018), and requires a larger number of drivers than customers (Mittal et al., 2021). Hence, driver attrition is not necessarily compensated by newcomers, or at least is not sustainable as long-term strategy.

Given the nature of the driver attrition operationalization and the characteristics of the dataset, following prior literature (Azadegan, Patel, & Parida, 2013; De Vries, DeKoster, & Stam, 2016; Singh, Kemerer, & Ramasubbu, 2021), we investigate driver attrition using survival analysis. Survival analysis is adopted when considering the occurrence of a binary outcome variable (presence or absence of an individual) given the left censored (late entry) and/or the right censored (truncation of the study period) nature of the dataset (Bhattacharjee et al., 2007; Flynn, 2012; Singh et al., 2021). Survival analysis can determine the impact of covariates on the survival time, defined as the time interval between the start and follow-up for a subject until the event of interest occurs or until censored (Flynn, 2012).

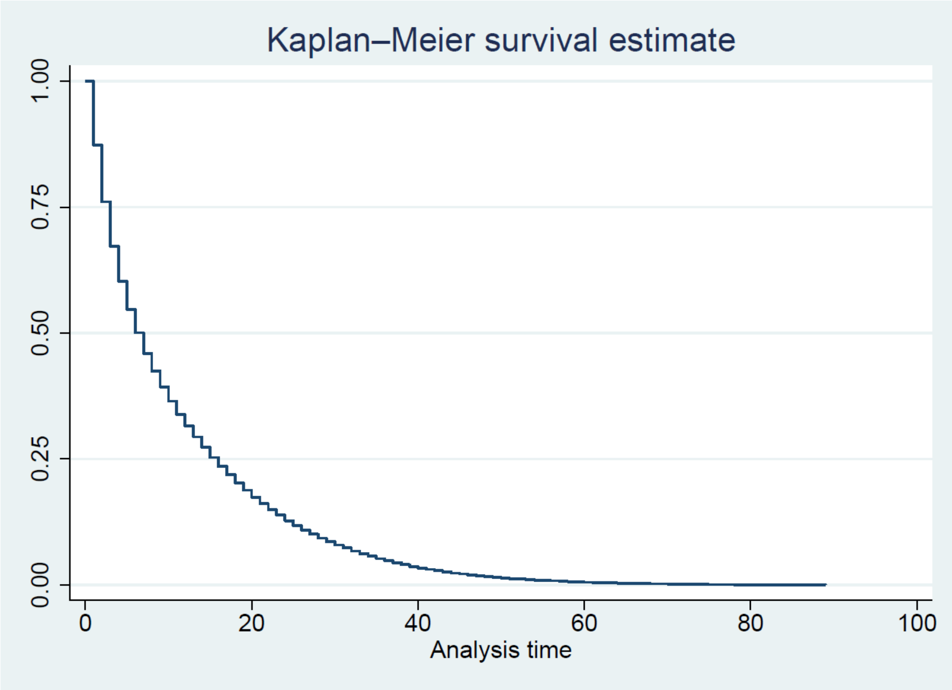


Figure 12 – Kaplan-Meier survival function curve

As common in survival analysis (Clark et al., 2003), we assess the survival rate of crowdshipping delivery drivers by consulting the Kaplan-Meier survival curve (Kaplan & Meier, 1958), which shows the survival function against time estimated as the cumulative probability of survival for all individuals in the dataset since the baseline (Kaplan & Meier, 1958). Thus, we identified the cumulative probability of how many drivers *survived* for how many days in the dataset before disappearing. Figure 12 presents the Kaplan-Meier curve showing that only 25% of delivery drivers (~19k drivers) in the dataset *survive* for more than 16-17 days. This *does not* reflect drivers that consistently delivered everyday for 17 consecutive days. Rather, it shows that only 25% of drivers were still present in the dataset after 17 days since their first appearance.

Motivated by this result and the results for driver return time, where we found that the most prominent predictors of retention are the monetary incentives, we assessed the impact of remuneration for a given day on the survival rate of crowdshipping delivery drivers. Following

prior literature (De Vries et al., 2018; Senot, 2019), we adopted several estimation models to assess the effect of covariates on the hazard rate and the robustness of the estimations (see Appendix A).

We adopted a Cox proportional hazards regression model (Cox, 1972). This is a semi-parametric multivariate survival analysis model, that estimates the effect of covariates on the hazard rate, without being subject to assumptions relative to the distribution of the hazard rate over time (Bradburn et al., 2003; Senot, 2019; Singh et al., 2021). Model 1 and Model 2 in Table 12 report the estimations for the impact of remuneration on the hazard rate without and with control variables by assessing the following estimation models:

$$h_d(t/\mathrm{X}) = \lambda_0(\mathrm{t}) \times \exp\{\alpha_1 M_{dy}\} \tag{15}$$

$$h_d(t/\mathrm{X}) = \lambda_0(\mathrm{t}) \times \exp\{\alpha_1 M_{dy} + \alpha_2 exp_{dy} + \alpha_3 ntask_{dy} + \alpha_4 ndr_y + \alpha_{4ym} month_{ym} + \alpha_{5yw} dow_{yw} + \alpha_{6dyt} time_{dyt}\} \tag{16}$$

Table 12 – Cox proportional hazard regression model for remuneration per day

	(1) $DRAT_{dy}$ β	se	(2) $DRAT_{dy}$ β	se
exp_{dy}			0.011^{****}	(0.000)
$ntask_{dy}$			-0.118^{****}	(0.005)
ndr_y			0.001^{****}	(0.000)
Month fe			YES	
Day of the week fe			YES	
Time of the day fe			YES	
Remuneration day	-0.004^{****}	(0.000)	-0.000^{+}	(0.000)
χ^2	4187.970		80078.372	
McFadden's pseudo R^2	0.003		0.05	
N	862,199		862,199	
N clusters	77,561		77,561	

Note: $DRAT_{dy}$ is driver attrition. $^{+}$ $p < 0.1$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.005$, **** $p < 0.001$. Reported robust standard errors are clustered on driver ID.

Next, Figure 13 report the Cox proportional hazards regression functions plotted at $\pm$ one standard deviation from the mean for each focal predictor. Specifically, in Figure 13, the curves represent the survival rate when the remuneration per day is \$23 (green), \$83 (red), and \$144 (blue). Interestingly, using the same 17 days threshold, Figure 13 suggests that plus one standard

deviation from the mean corresponds to an increase of the percentage of drivers from 25% (~19k drivers) to 35% (~27k drivers). Minus one standard deviation, instead, reduces drivers from 25% to 15% (~12k drivers).

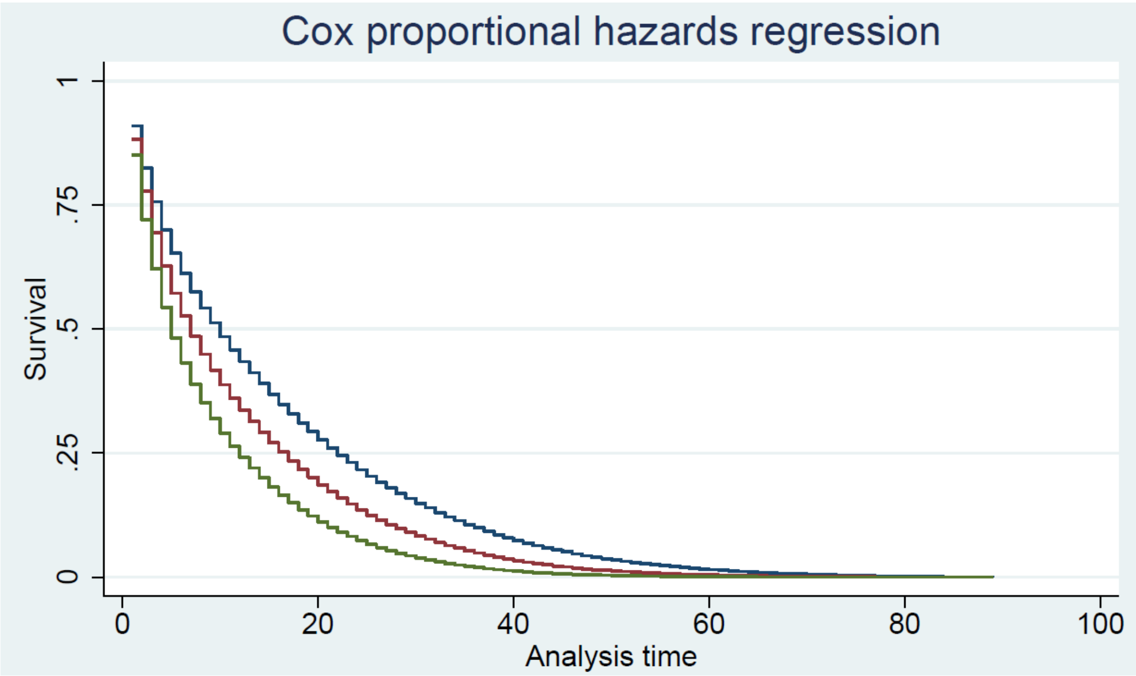


Figure 13 – Cox proportional hazard regression model for remuneration per day

Discussion of results for the effects on post-task behaviors – Driver satisfaction and retention

This set of results suggests that monetary incentives, rather than operational characteristics of the task, constitute the primary driver of post-task behaviors. The results for driver retention extend prior literature in several ways. First, this study contributes to the crowdshipping literature investigating driver retention through driver motivations (Anderson, 2014; Rosenblat, 2016; Cameron, 2022) and organizational identity (Petriglieri et al., 2019; Hua et al., 2020; Ai et al., 2023). We extend this literature by discussing the key role of the platform in providing a consistent flow of delivery tasks and income to delivery drivers, as well as the role of customers in providing the incentive to reduce attrition rate. Overall, greater remuneration does not guarantee that drivers will drive more consistently, but they tend to be more loyalty, whereas drivers will return more

frequently and are loyal to ensure additional tips. Second, the latter also expands the discussion about network effects in crowdshipping. Greater remuneration would attract more drivers, which would attract more customers, but it does not necessarily ease the matching and attract more customers, confirming that network effects is not always scalable due to the need for a large number of drivers (Qi et al., 2018; Cullen & Farronato, 2021; Mittal et al., 2021).

This set of results also contributes to theory. IOE identifies the effect of the income opportunity on the short-term performance of the server. Kamalahmadi et al. (2021) suggested investigating the impact of the income opportunity on the long-term impact on servers' turnover. We extend this perspective by finding the long-term (please recall that in the crowdshipping context three months is fairly long-term time period) consequences of the income opportunity. A consistent flow of income opportunities decreases server turnover. This is important for platforms that invest resources to nurture the relationship with the independent service provider (Apte & Davis, 2019), and fight issues such as lack of organizational identity (Ai et al., 2023). In addition, (Kamalahmadi et al., 2021) recommended studying the effect of the timing of the delivery task on a worker's response. We found that the delivery type does not offer a meaningful change in driver satisfaction and retention, extending the IOE by finding that the delivery characteristics do not significantly impact driver post-task behaviors.

Conclusions, limitations, and directions for future research

While crowdshipping is a growing phenomenon presenting several advantaged for last mile delivery, it brings challenges related to the uncertainty in the delivery driver supply capacity (Benjaafar & Hu, 2020). This study attempts to address such a challenge by investigating the role of monetary incentives and operational characteristics of the delivery task on driver behaviors in

the pre-task, task, and post-task. Besides the contributions presented after each set of results, we present here some overarching conclusions and contributions to theory and practice.

Theoretical contributions

This study presents two overarching theoretical contributions. First, the key importance of operational task characteristics in shaping driver behaviors in pre-task, task, and post-task. Using a social exchange perspective, a delivery driver would not just evaluate the monetary reward, but also the effort that will require to obtain such reward. This evaluation depends on the operational characteristics that the driver can notice before accepting the task (density, delivery type). Relatedly, some behaviors are influenced in a larger part by operational characteristics than monetary incentive. For instance, the Service time is more influenced by density and task complexity than remuneration. Hence, we draw research attention to consider monetary incentives as a minor source of motivation to improve the Service time (and quality) when compared to the operational characteristics. This, in turn, offers a contribution to IOE. The income opportunity may initially attract servers, but the service performance could be driven by motivations other than the income opportunity, such as the characteristics of the task. Finally, we provide empirical evidence to the IOE extension of SET by looking at the performance of the service provider.

Managerial contributions

This study also offers several managerial contributions. First, tracing driver behaviors throughout the delivery task offers the key advantage of understanding the effect of incentives on the driver performance. Typically, crowdshipping platforms present drivers with ex-post metrics relative to their performance. In contexts in which every minute counts, it would be possible to facilitate the service provision of an order, thus offering a better level of service to customers, by balancing the monetary incentives and the operational characteristics of the task. For example, combining orders

together to increase density can significantly reduce acceptance response time by 5 minutes, which corresponds to the Service time required to deliver an extra order. Importantly, it is the combined effect of monetary incentives and operational characteristics that attract a driver, and not just the greater remuneration. Similarly, a denser task can decrease the Service time by up to 10 minutes, which corresponds to twice the total time saving of the combined effect of remuneration and density for acceptance response time.

The second contribution to practice relates to service uncertainty and server capacity, which are limited by the intense competition between platforms (Bernstein et al., 2021) and high attrition rates (Rosenblat, 2016; Cook et al., 2021). This study informs managers that monetary incentives alone do not provide sufficient motivation to attract drivers, who might be loyal but not reliable for an on-demand service provision. Attracting drivers leveraging operational characteristics helps to reduce uncertainty as well as win over competitors. Finally, it is important to notice the key importance of batching customers' orders together in delivery tasks. Delivery drivers do not have immediate access to how much each individual customer tipped, rather, the pre-tipping disclosed is the sum of the pre-tipping for all orders. This is important, for example, to avoid driver retaliatory behaviors against a customers offering a low pre-tipping, as well as drivers deviating from the optimal routing (Liu et al., 2021) to satisfy customers offering larger pre-tipping.

Limitations and directions for future research

As with every empirical study, this research presents limitations. First, our study builds on the reasoning that three operational characteristics of the delivery task affect the relationship between monetary incentives and drivers behaviors. While we drew those operational characteristics from prior literature (Castillo et al., 2022b), and are theoretically and contextually interesting, there can

be other operational characteristics of the delivery task that we did not consider in this study. Future research may focus on the delivery task dispatching process flow, assessing how the interaction between store associates and drivers impacts the overall performance. Relatedly, we acknowledge the limitations of the dataset. First, the dataset includes only completed delivery tasks. However, a rising problem in last mile delivery are failed deliveries, which would require an expensive second delivery attempt (Cárdenas, Beckers, & Vanelslander, 2017; Convey, 2018). Studies could investigate how platforms handle such failed deliveries and if this aspect impact driver behaviors. Second, the operationalization of task complexity presents limitations. Future research could retrieve more granular data to explore the effect of scanning the barcodes of packages at the drop-off. Third, the dataset and the analyses only allows us to draw inference on driver behaviors, because we did not directly observed such behaviors through, for example, an ethnography (Castillo et al., 2022b). Future research could investigate such behaviors by adding qualitative data to the estimations results. Next, we based our study on a dataset retrieved from a single firm over a three-month period. While this limitation affects several studies in the operational management literature, a single-firm-dataset and from a quarter without peaks of seasonal delivery service demand (e.g., festivities) allows to obtain cleaner results (Lu, Lee, & Son, 2022). Hence, we encourage researchers and practitioners to cautiously interpret the conclusions of this study. In addition, the dataset may present limitations in terms of driver demographics, which have been found to be important predictors of retention, for example, gender (Cook et al., 2021). Finally, we encourage future research to expand the theoretical background to assess the impact of monetary and non-monetary incentives on driver behaviors.

Reference list

Aalen, O. O. 1989. A linear regression model for the analysis of life times. *Statistics in Medicine*, 8(8): 907–925.

Abbey, J. D., Meloy, M. G., Guide, V. D. R., & Atalay, S. 2015. Remanufactured products in closed-loop supply chains for consumer goods. *Production and Operations Management*, 24(3): 488–503.

Agatz, N., Fan, Y., & Stam, D. 2021. The Impact of Green Labels on Time Slot Choice and Operational Sustainability. *Production and Operations Management*, 30(7): 2285–2303.

Ai, W., Chen, Y., Mei, Q., Ye, J., & Zhang, L. 2023. Putting Teams into the Gig Economy: A Field Experiment at a Ride-Sharing Platform. *Management Science*, (February): 1–43.

Aiken, L. S., West, S. G., & Reno, R. R. 1991. *Multiple regression: Testing and interpreting interactions*. London: Sage.

Akerlof, G. A. 1982. Labor Contracts as Partial Gift Exchange. *The Quarterly Journal of Economics*, 97(4): 543.

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

Alexander, D., Boone, C., & Lynn, M. 2021. The effects of tip recommendations on customer tipping, satisfaction, repatronage, and spending. *Management Science*, 67(1): 146–165.

Amengual, M., & Apfelbaum, E. P. 2021. True motives: Prosocial and instrumental justifications for behavioral change in organizations. *Management Science*, 67(8): 5032–5051.

Amorim, P., DeHoratius, N., Eng-Larsson, F., & Martins, S. 2020. Customer Preferences for Delivery Service Attributes in Attended Home Delivery. *SSRN Electronic Journal*, (20).

Anderson, D. N. 2014. "Not Just a Taxi”? For-Profit Ridesharing, Driver Strategies, and VMT. *Transportation*, 41(5): 1099–1117.

Anderson, S., Allen, J., & Browne, M. 2005. Urban logistics - How can it meet policy makers’ sustainability objectives? *Journal of Transport Geography*, 13(1): 71–81.

Andritsos, D. A., & Tang, C. S. 2014. Linking process quality and resource usage: An empirical analysis. *Production and Operations Management*, 23(12): 2163–2177.

Apte, U. M., & Davis, M. M. 2019. Sharing economy services: Business model generation. *California Management Review*, 61(2): 104–131.

Arslan, A. M., Agatz, N., Kroon, L., & Zuidwijk, R. 2019. Crowdsourced delivery—a dynamic pickup and delivery problem with ad hoc drivers. *Transportation Science*, 53(1): 222–235.

Avgoustaki, A., & Bessa, I. 2019. Examining the link between flexible working arrangement bundles and employee work effort. *Human Resource Management*, 58(4): 431–449.

Awaysheh, A., Frohlich, M. T., Flynn, B. B., & Flynn, P. J. 2021. To err is human: Exploratory

multilevel analysis of supply chain delivery delays. *Journal of Operations Management*, 67(7): 882–916.

Azadegan, A., Patel, P. C., & Parida, V. 2013. Operational slack and venture survival. *Production and Operations Management*, 22(1): 1–18.

Banerjee, S., Freund, D., & Lykouris, T. 2022. Pricing and Optimization in Shared Vehicle Systems: An Approximation Framework. *Operations Research*, 70(3): 1783–1805.

Barker, J. M., Hofer, C., Hoberg, K., & Eroglu, C. 2022. Supplier inventory leanness and financial performance. *Journal of Operations Management*, 68(4): 385–407.

Barrios, J. M., Hochberg, Y. V., & Yi, H. 2022. The cost of convenience: Ridehailing and traffic fatalities. *Journal of Operations Management*, (September 2021): 1–33.

Batt, R. J., Diwas S., K. C., Staats, B. R., & Patterson, B. W. 2019. The Effects of Discrete Work Shifts on a Nonterminating Service System. *Production and Operations Management*, 28(6): 1528–1544.

Batt, R. J., & Gallino, S. 2019. Finding a needle in a haystack: The effects of searching and learning on pick-worker performance. *Management Science*, 65(6): 2624–2645.

Belenzon, S., Chatterji, A. K., & Daley, B. 2020. Choosing between growth and glory. *Management Science*, 66(5): 2050–2074.

Belo, R., Ferreira, P., & Telang, R. 2014. Broadband in school: Impact on student performance. *Management Science*, 60(2): 265–282.

Benjaafar, S., & Hu, M. 2020. Operations management in the age of the sharing economy: What is old and what is new? *Manufacturing and Service Operations Management*, 22(1): 93–101.

Benjamin, D. J., Berger, J. O., Johannesson, M., Nosek, B. A., Wagenmakers, E. J., Berk, R., Bollen, K. A., Brembs, B., Brown, L., Camerer, C., Cesarini, D., Chambers, C. D., Clyde, M., Cook, T. D., De Boeck, P., Dienes, Z., Dreber, A., Easwaran, K., Efferson, C., Fehr, E., et al. 2018. Redefine statistical significance. *Nature Human Behaviour*, 2(1): 6–10.

Benson, A., Sojourner, A., & Umyarov, A. 2020. Can reputation discipline the gig economy? Experimental evidence from an online labor market. *Management Science*, 66(5): 1802–1825.

Bernstein, F., DeCroix, G. A., & Keskin, N. B. 2021. Competition between two-sided platforms under demand and supply congestion effects. *Manufacturing and Service Operations Management*, 23(5): 1043–1061.

Bhattacharjee, S., Gopal, R. D., Lertwachara, K., Marsden, J. R., & Telang, R. 2007. The effect of digital sharing technologies on music markets: A survival analysis of albums on ranking charts. *Management Science*, 53(9): 1359–1374.

Boyer, K. K., Prud'homme, A. M., & Chung, W. 2009. The Last Mile Challenge: Evaluating the Effects Of Customer Density and Delivery Window Patterns. *Journal of Business Logistics*, 30(1): 185–201.

Bradburn, M. J., Clark, T. G., Love, S. B., & Altman, D. G. 2003. Survival Analysis Part II:

Multivariate data analysis- An introduction to concepts and methods. *British Journal of Cancer*, 89(3): 431–436.

Browning, T. R. 2020. Operations management writ large. *Journal of Operations Management*, 66(5): 494–500.

Cachon, G. P., Daniels, K. M., & Lobel, R. 2017. The Role of Surge Pricing on a Service Platform with Self-Scheduling Capacity. *Manufacturing & Service Operations Management*, 19(3): 368–384.

Cachon, G. P., Gallino, S., & Olivares, M. 2019. Does adding inventory increase sales? Evidence of a scarcity effect in U.S. Automobile dealerships. *Management Science*, 65(4): 1469–1485.

Cameron, L. D. 2022. "Making Out" While Driving: Relational and Efficiency Games in the Gig Economy. *Organization Science*, 33(1): 231–252.

Campbell, A. M., & Savelsbergh, M. W. P. 2005. Decision support for consumer direct grocery initiatives. *Transportation Science*, 39(3): 313–327.

Cantor, D. E., Macdonald, J. R., & Crum, M. R. 2011. The influence of workplace justice perceptions on commercial driver turnover intentions. *Journal of Business Logistics*, 32(3): 274–286.

Carbone, V., Rouquet, A., & Roussat, C. 2017. The Rise of Crowd Logistics: A New Way to Co-Create Logistics Value. *Journal of Business Logistics*, 38(4): 238–252.

Cárdenas, I., Beckers, J., & Vanelslander, T. 2017. E-commerce last-mile in Belgium: Developing an external cost delivery index. *Research in Transportation Business and Management*, 24(January): 123–129.

Castillo, V. E., Bell, J. E., Mollenkopf, D. A., & Stank, T. P. 2022a. Hybrid last mile delivery fleets with crowdsourcing: A systems view of managing the cost-service trade-off. *Journal of Business Logistics*, 43(1): 36–61.

Castillo, V. E., Bell, J. E., Rose, W. J., & Rodrigues, A. M. 2018. Crowdsourcing Last Mile Delivery: Strategic Implications and Future Research Directions. *Journal of Business Logistics*, 39(1): 7–25.

Castillo, V. E., Mollenkopf, D. A., Bell, J. E., & Esper, T. L. 2022b. Designing technology for on-demand delivery: The effect of customer tipping on crowdsourced driver behavior and last mile performance. *Journal of Operations Management*, 68(5): 424–453.

Cerasoli, C. P., Nicklin, J. M., & Ford, M. T. 2014. Intrinsic motivation and extrinsic incentives jointly predict performance: A 40-year meta-analysis. *Psychological Bulletin*, 140(4): 980–1008.

Certo, S. T., Busenbark, J. R., Woo, H., & Semadeni, M. 2016. Sample selection bias and Heckman models in strategic management research. *Strategic Management Journal*, 37(13): 2639–2657.

Certo, S. T., Withers, M. C., & Semadeni, M. 2017. A tale of two effects: Using longitudinal data to compare within- and between-firm effects. *Strategic Management Journal*, 38(7): 1536–

1556.

Chan, K. W., Li, S. Y., Ni, J., & Zhu, J. J. J. 2021. What Feedback Matters? The Role of Experience in Motivating Crowdsourcing Innovation. *Production and Operations Management*, 30(1): 103–126.

Chen, Y., & Wang, L. 2019. Commentary: Marketing and the Sharing Economy: Digital Economy and Emerging Market Challenges. *Journal of Marketing*, 83(5): 28–31.

Choudhary, V., Shunko, M., Netessine, S., & Koo, S. 2021. Nudging Drivers to Safety: Evidence from a Field Experiment. *Management Science*.

Chuang, H. H. C., Oliva, R., & Perdikaki, O. 2016. Traffic-Based Labor Planning in Retail Stores. *Production and Operations Management*, 25(1): 96–113.

Ciobotaru, G., & Chankov, S. 2021. Towards a taxonomy of crowdsourced delivery business models. *International Journal of Physical Distribution and Logistics Management*, 51(5): 460–485.

Clark, T. G., Bradburn, M. J., Love, S. B., & Altman, D. G. 2003. Survival Analysis Part I: Basic concepts and first analyses. *British Journal of Cancer*, 89(2): 232–238.

Convey. 2018. *Last Mile Delivery: What Shoppers Want and How to #SaveRetail*. *Convey*. Retrieved from https://www.getconvey.com/press-d-last-mile-delivery-save-retail/

Cook, C., Diamond, R., Hall, J. V., List, J. A., & Oyer, P. 2021. The gender earnings gap in the gig economy: Evidence from over a million rideshare drivers. *Review of Economic Studies*, 88(5): 2210–2238.

Cox, D. R. 1972. Regression Models and Life-Tables. *Journal of the Royal Statistical Society: Series B (Methodological)*, 34(2): 187–202.

Cropanzano, R., & Mitchell, M. S. 2005. Social exchange theory: An Interdisciplinary review. *Journal of Management*, 31(6): 874–900.

CSCMP. 2022. *33rd Annual State of Logistics Report*. *Council of supply chain management professionals*.

Cui, Y., & Davis, A. M. 2022. Tax-Induced Inequalities in the Sharing Economy. *Management Science*, (May).

Cullen, Z., & Farronato, C. 2021. Outsourcing tasks online: Matching supply and demand on peer-to-peer internet platforms. *Management Science*, 67(7): 3985–4003.

Cunningham, C., Gerardi, K., & Shen, L. 2021. The double trigger for mortgage default: Evidence from the fracking boom. *Management Science*, 67(6): 3943–3964.

Dai, H., & Liu, P. 2020. Workforce planning for O2O delivery systems with crowdsourced drivers. *Annals of Operations Research*, 291(1–2): 219–245.

Das, T. K., & Teng, B.-S. 2002. Alliance Constellations: A Social Exchange Perspective. *Academy of Management Review*, 27(3): 445–456.

Davis, G. F. 2016. Can an economy survive without corporations? Technology and robust

organizational alternatives. *Academy of Management Perspectives*, 30(2): 129–140.

Dayarian, I., & Pazour, J. 2022. Crowdsourced order-fulfillment policies using in-store customers. *Production and Operations Management*, 1–32.

Dayarian, I., & Savelsbergh, M. 2020. Crowdshipping and Same-day Delivery: Employing In-store Customers to Deliver Online Orders. *Production and Operations Management*, 29(9): 2153–2174.

De Vries, J., DeKoster, R., & Stam, D. 2016. Aligning Order Picking Methods, Incentive Systems, and Regulatory Focus to Increase Performance. *Production and Operations Management*, 25(8): 1363–1376.

De Vries, J., Roy, D., & De Koster, R. 2018. Worth the wait? How restaurant waiting time influences customer behavior and revenue. *Journal of Operations Management*, 63(May): 59–78.

Delfgaauw, J., Dur, R., Onemu, O., & Sol, J. 2022. Team Incentives, Social Cohesion, and Performance: A Natural Field Experiment. *Management Science*, 68(1): 230–256.

Devari, A., Nikolaev, A. G., & He, Q. 2017. Crowdsourcing the last mile delivery of online orders by exploiting the social networks of retail store customers. *Transportation Research Part E: Logistics and Transportation Review*, 105: 105–122.

Dhanorkar, S., & Siemsen, E. 2021. How Nudges Lead to Improved Energy Efficiency in Manufacturing: Evidence from Archival Data and a Field Study. *Production and Operations Management*, 30(10): 3735–3757.

Dhanorkara, S. 2019. Environmental benefits of internet-enabled c2c closed-loop supply chains: A quasi-experimental study of craigslist. *Management Science*, 65(2): 660–680.

Dikolli, S. S., Heater, J. C., Mayew, W. J., & Sethuraman, M. 2021. Chief financial officer co-option and chief executive officer compensation. *Management Science*, 67(3): 1939–1955.

Diwas, S. K. C., Staats, B. R., Kouchaki, M., & Gino, F. 2020. Task selection and workload: A focus on completing easy tasks hurts performance. *Management Science*, 66(10): 4397–4416.

Dixon, J., Hong, B., & Wu, L. 2021. The robot revolution: Managerial and employment consequences for firms. *Management Science*, 67(9): 5586–5605.

Dong, J., & Ibrahim, R. 2020. Managing supply in the on-demand economy: Flexible workers, full-time employees, or both? *Operations Research*, 68(4): 1238–1264.

Douglas, M. 2020. Winning Moves for the Last Mile. *Inbound Logistics*. Retrieved February 26, 2020, from https://www.inboundlogistics.com/cms/article/winning-moves-for-the-last-mile/

Drake, M., Joos, P., Pacelli, J., & Twedt, B. 2020. Analyst forecast bundling. *Management Science*, 66(9): 4024–4046.

Duhaime, E. P., & Woessner, Z. W. 2019. Explaining the decline of tipping norms in the gig economy. *Journal of Managerial Psychology*, 34(4): 233–245.

Eckhardt, G. M., Houston, M. B., Jiang, B., Lamberton, C., Rindfleisch, A., & Zervas, G. 2019.

Marketing in the Sharing Economy. *Journal of Marketing*, 83(5): 5–27.

Emadi, S. M., & Staats, B. R. 2020. A structural estimation approach to study agent attrition. *Management Science*, 66(9): 4071–4095.

Emerson, R. M. 1962. Power-Dependence Relations. *American Sociological Review*, 27(1): 31–41.

Ergün-Şahin, B., Güneş, E. D., Kocabıyıkoğlu, A., & Keskin, A. 2022. How does workload affect test ordering behavior of physicians? An empirical investigation. *Production and Operations Management*, (Forthcoming).

Esper, T. L., Castillo, V. E., Ren, K., Sodero, A., Wan, X., Croxton, K. L., Knemeyer, A. M., DeNunzio, S., Zinn, W., & Goldsby, T. J. 2020. Everything Old is New Again: The Age of Consumer-Centric Supply Chain Management. *Journal of Business Logistics*, 41(4): 286–293.

Farber, H. S. 2015. Why you can't find a taxi in the rain and other labor supply lessons from cab drivers. *Quarterly Journal of Economics*, 130(4): 1975–2026.

Fatehi, S., & Wagner, M. R. 2022. Crowdsourcing Last-Mile Deliveries. *Manufacturing & Service Operations Management*, 24(2): 791–809.

Filippas, A., Horton, J. J., & Zeckhauser, R. J. 2020. Owning, using, and renting: Some simple economics of the "sharing economy." *Management Science*, 66(9): 4152–4172.

Flynn, F. J. 2005. Identity orientations and forms of social exchange in organizations. *Academy of Management Review*, 30(4): 737–750.

Flynn, R. 2012. Survival analysis. *Journal of Clinical Nursing*, 21(19–20): 2789–2797.

Gambeta, E., Koka, B. R., & Hoskisson, R. E. 2019. Being too good for your own good: A stakeholder perspective on the differential effect of firm-employee relationships on innovation search. *Strategic Management Journal*, 40(1): 108–126.

Gao, F., & Su, X. 2017. Omnichannel retail operations with buy-online-and-pick-up-in-store. *Management Science*, 63(8): 2478–2492.

Garland, M. 2022. How Walmart GoLocal plans to stand out among delivery businesses. *Supply Chain Dive*. Retrieved February 21, 2022, from https://www.supplychaindive.com/news/walmart-golocal-new-delivery-business/617117/

Gerwe, O., & Silva, R. 2020. Clarifying the sharing economy: Conceptualization, typology, antecedents, and effects. *Academy of Management Perspectives*, 34(1): 65–96.

Griffith, D. A., Harvey, M. G., & Lusch, R. F. 2006. Social exchange in supply chain relationships: The resulting benefits of procedural and distributive justice. *Journal of Operations Management*, 24(2): 85–98.

Guda, H., & Subramaniana, U. 2019. Your uber is arriving: Managing on-demand workers through surge pricing, forecast communication, and worker incentives. *Management Science*, 65(5): 1995–2014.

Gümüsay, A. A. 2018. Commentary: Sharing is Caring: From Material to Socio-Material Sharing. *Academy of Management Discoveries*, 4(3): 376–377.

Hasija, S., Shen, Z. J. M., & Teo, C. P. 2020. Smart city operations: Modeling challenges and opportunities. *Manufacturing and Service Operations Management*, 22(1): 203–213.

Hausman, J. A. 1978. Specification Tests in Econometrics. *Econometrica*, 46(6): 1251–1271. Retrieved from http://www.jstor.org/stable/1913827%5Cnhttp://www.jstor.org/%5Cnhttp://www.jstor.org/action/showPublisher?publisherCode=econosoc.%5Cnhttp://www.jstor.org

Heckman, J. J. 1979. Sample Selection Bias as a Specification Error. *Econometrica*, 47(1): 153.

Ho, T. H., Lim, N., Reza, S., & Xia, X. 2017. OM forum - Causal inference models in operations management. *Manufacturing and Service Operations Management*, 19(4): 509–525.

Holm, D. B., Eriksson, K., & Johanson, J. 1999. Creating value through mutual commitment to business network relationships. *Strategic Management Journal*, 20(5): 467–486.

Homans, G. C. 1958. Social Behavior as Exchange. *American Journal of Sociology*, 63(6): 597–606.

Hopp, W. J., & Simchi-Levi, D. 2021. Management science: The legacy of the past and challenge of the future. *Management Science*, 67(9): 5306–5316.

Hosmer, D., & Lemeshow, S. 1999. *Applied survival analysis: regression modeling of time to event data*. New York: JohnWiley and sons.

Howe, M., & Jin, Y. 2022. It's nothing personal, or is it? Exploring the competitive implications of relational multiplexity in supply chains. *Journal of Supply Chain Management*, 58(2): 26–47.

Hu, M. 2021. From the Classics to New Tunes: A Neoclassical View on Sharing Economy and Innovative Marketplaces. *Production and Operations Management*, 30(6): 1668–1685.

Hua, Y., Cheng, X., Hou, T., & Luo, R. 2020. Monetary Rewards, Intrinsic Motivators, and Work Engagement in the IT-Enabled Sharing Economy: A Mixed-Methods Investigation of Internet Taxi Drivers*. *Decision Sciences*, 51(3): 755–785.

Huffman, D., & Bognanno, M. 2018. High-Powered Performance Pay and Crowding Out of Nonmonetary Motives. *Management Science*, 64(10): 4669–4680.

Ibrahim, R. 2018. Managing Queueing Systems Where Capacity is Random and Customers are Impatient. *Production and Operations Management*, 27(2): 234–250.

Ishfaq, R., & Raja, U. 2018. Evaluation of Order Fulfillment Options in Retail Supply Chains. *Decision Sciences*, 49(3): 487–521.

Islam, S., Uddin, M. J., Shi, Y., Sharif, T., & Ahmed, J. U. 2020. Exploring shippers' motivations to adopt collaborative truck-sharing initiatives. *International Journal of Physical Distribution and Logistics Management*, 51(5): 528–550.

Jain, N., & Tan, T. F. 2022. M-Commerce, Sales Concentration, and Inventory Management.

Manufacturing & Service Operations Management, 24(4): 2256–2273.

Janjevic, M., & Winkenbach, M. 2020. Characterizing urban last-mile distribution strategies in mature and emerging e-commerce markets. *Transportation Research Part A: Policy and Practice*, 133: 164–196.

Jerkings Jr., G. D., Mitra, A., Gupta, N., & Shaw, J. D. 1998. Are Financial Incentives Related to Performance? A Meta-Analytic Review of Empirical Research. *Journal of Applied Psychology*, 83(5): 777–787.

Kamalahmadi, M., Yu, Q., & Zhou, Y. P. 2021. Call to Duty: Just-in-Time Scheduling in a Restaurant Chain. *Management Science*, 67(11): 6751–6781.

Kaplan, E. L., & Meier, P. 1958. Nonparametric Estimation from Incomplete Observations. *Journal of the American Statistical Association*, 53(282): 457–481.

Kathan, W., Matzler, K., & Veider, V. 2016. The sharing economy: Your business model's friend or foe? *Business Horizons*, 59(6): 663–672.

KC, D. S., & Terwiesch, C. 2009. Impact of workload on service time and patient safety: An econometric analysis of hospital operations. *Management Science*, 55(9): 1486–1498.

Kennedy, P. 2008. *A guide to econometrics*. Cambridge, MA: John Wiley & Sons.

Kim, D. Y., & Zhu, P. 2018. Supplier dependence and R&D intensity: The moderating role of network centrality and interconnectedness. *Journal of Operations Management*, 64(September 2017): 7–18.

Köbis, C. N., Soraperra, I., & Shalvi, S. 2021. The Consequences of Participating in the Sharing Economy: A Transparency-Based Sharing Framework. *Journal of Management*, 47(1): 317–343.

Kooti, F., Djuric, N., Grbovic, M., Radosavljevic, V., Aiello, L. M., & Lerman, K. 2019. Analyzing uber's ride-sharing economy. *26th International World Wide Web Conference 2017, WWW 2017 Companion*, (January): 574–582.

Kornberger, M., Leixnering, S., Meyer, R. E., & Höllerer, M. A. 2018. Rethinking the Sharing Economy: The Nature and Organization of Sharing in the 2015 Refugee Crisis. *Academy of Management Discoveries*, 4(3): 314–335.

Lambe, C. J., Wittmann, C. M., & Spekman, R. E. 2001. Social exchange theory and research on business-to-business relational exchange. *Journal of Business-to-Business Marketing*, 8(3): 1–36.

Lehdonvirta, V., Kässi, O., Hjorth, I., Barnard, H., & Graham, M. 2019. The Global Platform Economy: A New Offshoring Institution Enabling Emerging-Economy Microproviders. *Journal of Management*, 45(2): 567–599.

Li, Z., Liang, C., Hong, Y., & Zhang, Z. 2022. How Do On-demand Ridesharing Services Affect Traffic Congestion? The Moderating Role of Urban Compactness. *Production and Operations Management*, 31(1): 239–258.

Liao, H., Toya, K., Lepak, D. P., & Hong, Y. 2009. Do They See Eye to Eye? Management and

Employee Perspectives of High-Performance Work Systems and Influence Processes on Service Quality. *Journal of Applied Psychology*, 94(2): 371–391.

Lim, S. F. W. T., Jin, X., & Srai, J. S. 2018. Consumer-driven e-commerce: A literature review, design framework, and research agenda on last-mile logistics models. *International Journal of Physical Distribution and Logistics Management*, 48(3): 308–332.

Lim, S. F. W. T., Wang, Q., & Webster, S. 2023. Do it right the first time: Vehicle routing with home delivery attempt predictors. *Production and Operations Management*, Forthcomin.

Lin, M., Lucas, H. C., & Shmueli, G. 2013. Too big to fail: Large samples and the p-value problem. *Information Systems Research*, 24(4): 906–917.

Liu, H., Wei, S., Ke, W., Wei, K. K., & Hua, Z. 2016. The configuration between supply chain integration and information technology competency: A resource orchestration perspective. *Journal of Operations Management*, 44: 13–29.

Liu, S., He, L., & Shen, Z. J. M. 2021. On-time last-mile delivery: Order assignment with travel-time predictors. *Management Science*, 67(7): 4095–4119.

Liu, Y., Luo, Y., & Liu, T. 2009. Governing buyer-supplier relationships through transactional and relational mechanisms: Evidence from China. *Journal of Operations Management*, 27(4): 294–309.

Lu, G., Ding, X. (David), Peng, D. X., & Hao-Chun Chuang, H. 2018. Addressing endogeneity in operations management research: Recent developments, common problems, and directions for future research. *Journal of Operations Management*, 64(1): 53–64.

Lu, G., Lee, H. S., & Son, J. 2022. Product variety in local grocery stores: Differential effects on stock-keeping unit level sales. *Journal of Operations Management*, 68(1): 33–54.

Lu, G., & Shang, G. 2017. Impact of supply base structural complexity on financial performance: Roles of visible and not-so-visible characteristics. *Journal of Operations Management*, 53–56(March 2016): 23–44.

Lumineau, F., & Henderson, J. E. 2012. The influence of relational experience and contractual governance on the negotiation strategy in buyer-supplier disputes. *Journal of Operations Management*, 30(5): 382–395.

Mak, H. Y. 2022. Enabling Smarter Cities with Operations Management. *Manufacturing and Service Operations Management*, 24(1): 24–39.

Mao, W., Ming, L., Rong, Y., Tang, C. S., & Zheng, H. 2019. Faster Deliveries and Smarter Order Assignments for an On-Demand Meal Delivery Platform. *SSRN Electronic Journal*.

Mayo, A. 2021. Amazon delivery drivers say there's a "giant war" between them and the company as they struggle to meet package quotas. *Business Insider*. Retrieved November 6, 2022, from https://www.businessinsider.com/amazon-drivers-interview-giant-war-between-them-and-company-packages-2021-7

Miao, W., Deng, Y., Wang, W., Liu, Y., & Tang, C. S. 2022. The effects of surge pricing on driver behavior in the ride-sharing market: Evidence from a quasi-experiment. *Journal of*

Operations Management, Forthcomin.

Mitchell, T. R., & Mickel, A. E. 1999. The Meaning of Money: An Individual-Difference Perspective. *Academy of Management Review*, 24(3): 568–578.

Mithas, S., Chen, Y., Lin, Y., & De Oliveira Silveira, A. 2022. On the causality and plausibility of treatment effects in operations management research. *Production and Operations Management*, 31(12): 4558–4571.

Mittal, A., Oran Gibson, N., Krejci, C. C., & Marusak, A. A. 2021. Crowd-shipping for urban food rescue logistics. *International Journal of Physical Distribution and Logistics Management*, 51(5): 486–507.

Mohajeri, K., Mesgari, M., & Lee, A. S. 2020. When statistical significance is not enough: Investigating relevance, practical significance, and statistical significance. *MIS Quarterly: Management Information Systems*, 44(2): 525–559.

Mulvey, J. M. 1983. Multivariate Stratified Sampling by Optimization. *Management Science*, 29(6): 715–724.

Muñoz, P., & Cohen, B. 2018. A compass for navigating sharing economy business models. *California Management Review*, 61(1): 114–147.

Na, H. S., Kweon, S. J., & Park, K. 2021. Characterization and Design for Last Mile Logistics: A Review of the State of the Art and Future Directions. *Applied Sciences*, 12(1): 118.

Narasimhan, R., Nair, A., Griffith, D. A., Arlbjørn, J. S., & Bendoly, E. 2009. Lock-in situations in supply chains: A social exchange theoretic study of sourcing arrangements in buyer-supplier relationships. *Journal of Operations Management*, 27(5): 374–389.

Patel, P. C., Baldauf, C., Karlsson, S., & Oghazi, P. 2021. The impact of free returns on online purchase behavior: Evidence from an intervention at an online retailer. *Journal of Operations Management*, 67(4): 511–555.

Peinkofer, S. T., Schwieterman, M. A., & Miller, J. W. 2020. Last-mile delivery in the motor-carrier industry: A panel data investigation using discrete time event history analysis. *Transportation Journal*, 59(3): 129–164.

Perdikaki, O., Peng, D. X., & Heim, G. R. 2015. Impact of Customer Traffic and Service Process Outsourcing Levels on e-Retailer Operational Performance. *Production and Operations Management*, 24(11): 1794–1811.

Petriglieri, G., Ashford, S. J., & Wrzesniewski, A. 2019. Agony and Ecstasy in the Gig Economy: Cultivating Holding Environments for Precarious and Personalized Work Identities. *Administrative Science Quarterly*, 64(1): 124–170.

Picard, R. 2022. help geodist. Retrieved from http://fmwww.bc.edu/repec/bocode/g/geodist.html

Qi, W., Li, L., Liu, S., & Shen, Z. M. 2018. Shared Mobility for Last-Mile Delivery: Design, Operational Prescriptions, and Environmental Impact. *Manufacturing & Service Operations Management*, 20(4): 737–751.

Qi, Y., Wang, X., Zhang, M., & Wang, Q. 2022. Developing supply chain resilience through

integration: An empirical study on an e-commerce platform. *Journal of Operations Management*, (October): 1–20.

Ramasubbu, N., & Kemerer, C. F. 2016. Technical debt and the reliability of enterprise software systems: A competing risks analysis. *Management Science*, 62(5): 1487–1510.

Rhea, C. M. 2022. Instacart Pay: How Much Does Instacart Pay Shoppers? *therideshareguy.com*. Retrieved from https://therideshareguy.com/instacart-shopper-pay/

Richey, R. G., & Davis-Sramek, B. 2022. A research agenda to reflect reality: On being responsive. *Journal of Business Logistics*, 43(1): 4–8.

Rosenblat, A. 2016. What Motivates Gig Economy Workers. *Harvard Business Review*, 2–5.

Salari, N., Liu, S., & Shen, Z.-J. M. 2022. Real-Time Delivery Time Forecasting and Promising in Online Retailing: When Will Your Package Arrive? *Manufacturing & Service Operations Management*, 24(3): 1421–1436.

Senot, C. 2019. Continuity of Care and Risk of Readmission: An Investigation into the Healthcare Journey of Heart Failure Patients. *Production and Operations Management*, 28(8): 2008–2030.

Shang, G., Ferguson, M. E., & Galbreth, M. R. 2019. Where Should I Focus My Return Reduction Efforts? Empirical Guidance for Retailers. *Decision Sciences*, 50(4): 877–909.

Shen, Z. J. M., & Daskin, M. S. 2005. Trade-offs between customer service and cost in integrated supply chain design. *Manufacturing and Service Operations Management*, 7(3): 188–207.

Singh, S., Kemerer, C. F., & Ramasubbu, N. 2021. Innovation in complex assembled electronic products: An analysis of the evolution of television components. *Journal of Operations Management*, 67(6): 680–703.

Song, H., Andreyeva, E., & David, G. 2022. Time Is the Wisest Counselor of All: The Value of Provider–Patient Engagement Length in Home Healthcare. *Management Science*, 68(1): 420–441.

Spector, P. E., & Fox, S. 2002. An emotion-centered model of voluntary work behavior. *Human Resource Management Review*, 12(2): 269–292.

Straightaway. 2022. UPS delivery route planner. Retrieved November 6, 2022, from https://www.getstraightaway.com/ups

Suk, I., Lee, S., & Kross, W. 2021. CEO turnover and accounting earnings: The role of earnings persistence. *Management Science*, 67(5): 3195–3218.

Ta, H., Esper, T. L., & Rossiter, A. 2018. Designing crowdsourced delivery systems : The effect of driver disclosure and ethnic similarity. *Journal of Operations Management*, 60(1): 19–33.

Ta, H., Esper, T. L., & Tokar, T. 2021. Appealing to the Crowd: Motivation Message Framing and Crowdsourcing Performance in Retail Operations. *Production and Operations Management*, 30(9): 3192–3212.

Taylor, T. A. 2018. On-Demand Service Platforms. *Manufacturing & Service Operations*

Management, 20(4): 704–720.

Tereyağoğlu, N., Fader, P. S., & Veeraraghavan, S. 2017. Pricing Theater Seats: The Value of Price Commitment and Monotone Discounting. *Production and Operations Management*, 26(6): 1056–1075.

Terpend, R., & Krause, D. R. 2015. Competition or Cooperation? Promoting Supplier Performance with Incentives Under Varying Conditions of Dependence. *Journal of Supply Chain Management*, 51(4): 29–53.

Thibaut, J. W., & Kelley, H. H. 1959. *The social psychology of groups*. New York: John Wiley & Sons.

Thibaut, J. W., & Kelly, H. H. 1959. Analysis and Concepts. *The Social Psychology of Groups* (pp. 9–30). New York: Wiley & Sons.

Tokar, T., & Swink, M. 2019. Public Policy and Supply Chain Management: Using Shared Foundational Principles to Improve Formulation, Implementation, and Evaluation. *Journal of Supply Chain Management*, 55(2): 68–79.

Tripathy, M., Bai, J., & Heese, H. S. 2022. Driver collusion in ride-hailing platforms. *Decision Sciences*, (March 2021): 1–13.

Vincenty, T. 1975. Direct and Inverse Solutions of Geodesics on the Ellipsoid with Application of Nested Equations. *Survey Review*, 23(176): 88–93.

Walmart.com. 2023. Spark Driver Service Level Standards - My Metrics. Retrieved April 3, 2023, from https://drive4spark.walmart.com/My Metrics

Wang, J., & Zhou, Y. P. 2018. Impact of queue configuration on service time: Evidence from a supermarket. *Management Science*, 64(7): 3055–3075.

Wang, L., Rabinovich, E., & Guda, H. 2022. An analysis of operating efficiency and policy implications in last-mile transportation following Amazon's integration. *Journal of Operations Management*, 1–27.

Wang, T., & Wu, D. 2020. Data-Driven Driver Dispatching System with Allocation Constraints and Operational Risk Management for a Ride-Sharing Platform. *Decision Sciences*, 51(6): 1490–1520.

Wang, Y., Zhang, D., Liu, Q., Shen, F., & Lee, L. H. 2016. Towards enhancing the last-mile delivery: An effective crowd-tasking model with scalable solutions. *Transportation Research Part E: Logistics and Transportation Review*, 93: 279–293.

Wei, Z., Xiao, M., & Rong, R. 2021. Network Size and Content Generation on Social Media Platforms. *Production and Operations Management*, 30(5): 1406–1426.

Wooldridge, J. M. 2010. *Econometric Analysis of Cross Section and Panel Data*. MIT press.

Wooldridge, J. M. 2016. *Introductory econometrics: A modern approach*. Mason, OH: Nelsen Education.

Yildiz, B., & Savelsbergh, M. 2019. Service and capacity planning in crowd-sourced delivery.

Transportation Research Part C: Emerging Technologies, 100(January): 177–199.

Yu, J. J., Tang, C. S., Shen, Z. J. M., & Chen, X. M. 2020. A balancing act of regulating on-demand ride services. *Management Science*, 66(7): 2975–2992.

Zebra Technologies. 2018. The Future of Fulfillment Vision Study. *Zebra*, 12. Retrieved from https://www.zebra.com/content/dam/zebra_new_ia/en-us/solutions-verticals/vertical-solutions/retail/vision-study/fulfillment-vision-study-report-en-us.pdf

Zhang, C., Kettinger, W. J., Kolte, P., & Yoo, S. 2018. Established companies' strategic responses to sharing economy threats. *MIS Quarterly Executive*, 17(1): 23–40.

III. Essay 2: Leveraging driver's learning: The impact of driver's familiarity and order characteristics on delivery performance

Introduction

Improving last mile delivery performance has become increasingly important to ensure service quality in the order fulfillment process (Shang & Liu, 2011; Caro, Kök, & Martínez-de-Albéniz, 2020; Cui, Li, & Li, 2020). However, delivery service providers struggle with on-time deliveries (Liu, He, & Shen, 2021). Recent empirical evidence and anecdotal statistics showcase the relevance of delivery performance for final customers and service providers (Convey, 2018; Cui, Yang, & Vertinsky, 2018; Awaysheh et al., 2021; Dolan, 2022). Prior studies examined several approaches to improve on-time performance. A stream of literature adopted analytical models to improve efficiency and effectiveness (Lu, Suzuki, & Clottey, 2020; Paradiso et al., 2020; Delasay, Jain, & Kumar, 2022; Stroh, Erera, & Toriello, 2022). An alternative approach was to explore the impact of innovation in last mile delivery on performance with studies on drone delivery, automatic vehicles, and real-time vehicle tracking (Perera et al., 2020; Praet & Martens, 2020; Reed, Campbell, & Thomas, 2022). Finally, literature has also looked at improving efficiency through drop-offs consolidation and enabling customers to self-delivery (Wollenburg et al., 2018; Wang et al., 2019a; Lyu & Teo, 2022).

An alternative solution is to understand delivery drivers' experience as a key resource. Specifically, to improve delivery performance, studies have looked at driver familiarity, defined as the experience of the driver repeatedly visiting service areas or a customer (Zhong, Hall, & Dessouky, 2007). Familiarity is a crucial factor influencing delivery performance, especially crowdshipping (Mao et al., 2019, 2022). Several studies advocate for assigning delivery tasks to drivers familiar with the service area or the final customer (Smilowitz, Nowak, & Jiang, 2013).

Driver familiarity provides the advantage of decreasing the service time while improving driver and customer outcomes (Keller, 2002; Haughton, 2008). Indeed, a driver that develops expertise through familiarity will improve operations in the service time, such as finding the destination, parking, and delivering (Schneider, 2016). However, a contrasting view indicates that driver familiarity is detrimental to delivery performance (Zhong et al., 2007). Increased familiarity implies fixed routes and territories that are repeatedly assigned to the same drivers. In the long run, this reduces flexibility in optimizing dispatch plans on a daily basis, and drivers would also perform worse when delivering in other areas or for other customers (Zhong et al., 2007). Planning flexibility is especially important in crowdshipping, where platforms need to match random demand with random supply (Fatehi & Wagner, 2022; Tian, Shi, & Qi, 2022). Hence, familiarity is valuable when drivers are dispatched to their delivery zone. This study supports the logic that driver familiarity improves delivery performance. This is consistent with the crowdshipping context, in which drivers are assigned to delivery zones (Cachon, Daniels, & Lobel, 2017). Indeed, assigning drivers to the same delivery zone or customers will have a learning effect that, over time, increases delivery time performance, customer satisfaction with the delivery, and driver retention (Zhong et al., 2007). Hence, the first research question is RQ1: *How does driver familiarity impact delivery outcomes in terms of time performance, customer satisfaction, and driver retention?*

The role of driver familiarity is likely to be contingent upon delivery characteristics. The first characteristics influencing driver familiarity with a delivery zone, or a customer relates to the type of delivery, specifically attended or unattended delivery. An attended delivery occurs when there is a service encounter between the customer and the delivery driver (McKinnon & Tallam, 2003; Lim, Jin, & Srai, 2018). Conversely, unattended deliveries are provided through reception boxes, and they usually improve efficiency (Punakivi, Yrjölä, & Holmström, 2001). In

crowdshipping, delivery drivers develop expertise in delivering either attended or unattended deliveries. Cameron (2022) explains that crowdshipping drivers play either a relational or efficiency game. In the relational game, drivers develop expertise in establishing a positive relationship with the customer. On the other hand, in the efficiency game, drivers develop expertise in improving delivery efficiency by avoiding the interaction with the customer. Hence, the extent to which drivers utilize either of the two delivery types will impact how familiarity affects delivery, customer, and driver outcomes. The second research question is RQ2: *How do delivery repeated characteristics affect the relationship between driver's familiarity and delivery outcomes in terms of time performance, customer satisfaction, and driver retention?*

A second element related to the delivery characteristics is the geographic area in which the delivery is completed. Literature on last mile delivery and crowdshipping has discussed the contextual differences between rural and urban delivery (Rose, Bell, & Griffis, 2022). Specifically, rural deliveries are characterized by less complexity because of the fewer constraints related, for example, to less traffic, time constraints, and density (Rose et al., 2019; Seghezzi, Siragusa, & Mangiaracina, 2022). In contrast, urban delivery requires more experienced drivers to overcome a large city's uncertainty, accessibility, and redundancy (Rose et al., 2019). Congestion and complexity of urban scenarios decrease the chances for drivers to become familiar with the delivery area, even though operations managers often assign a subset of experienced drivers to an urban core, while interchanging drivers among more peripheral areas (Rose et al., 2019; Seghezzi et al., 2022). This is also important because when drivers become more familiar with their routes over time, it should be possible to add more customers or expand the service territory (Smilowitz et al., 2013). This is consistent with the notion that as a worker completes more varied tasks, he/she can change more effectively between tasks (Staats & Gino, 2012). Hence, driver familiarity with

the delivery route is also contingent upon the delivery area. The third research question is RQ3: *How do delivery characteristics in term of delivery area affect the relationship between driver familiarity and delivery outcomes in terms of time performance, customer satisfaction, and driver's retention?*

To inform these research questions, this study adopts the logic of the body of knowledge on learning in service systems (Shafer, Nembhard, & Uzumeri, 2001; Staats & Gino, 2012; Bimpikis & Markakis, 2019). This body of knowledge theorizes on the interplay between service time and task-type uncertainty, defining the performance of two categories of servers: junior and senior (Bimpikis & Markakis, 2019). We complement these notions with recent literature on last mile delivery that explains how the contextual elements of crowdshipping play a role in driver familiarity. We develop hypotheses relative to the direct effect of driver familiarity on delivery outcomes and the moderation effect of delivery characteristics.

We empirically investigate these hypotheses using a large dataset comprising 7 million deliveries performed over a three-month period (February to April 2022). Results from econometric analysis offer several theoretical and managerial contributions. First, we reveal the role of driver familiarity in shaping delivery outcomes, adding knowledge to the debate in the literature between the positive and detrimental effects of driver familiarity on delivery performance. Second, we identify the boundary conditions under which learning improves performance, determining the impact of driver attitude toward deliveries, and the impact of geographic areas on learning. Finally, we aim to provide managerial implications relative to dispatching deliveries to experienced drivers and the relationship between driver and customer in B2C last mile delivery.

Literature review: Operational challenges in last mile delivery

The literature focused on optimizing the activities within last mile delivery by primarily adopting analytical models to address efficiency issues related to new challenges in last mile delivery. As demand for the last mile delivery service has increased, Paradiso et al. (2020) proposed an exact solution framework to solve multi-trip vehicle-routing in last mile delivery with instances up to 50 customers, as compared to previous models that could not consistently find an exact solution with more than 25 customers. Relatedly, Stroh et al. (2022) designed delivery systems planned to face time pressure for same-day delivery and uncertainty in level of order variability. Further, Hsiao et al. (2018) formulated a model to address time pressure related to short time windows and cold chain perishable items in last mile delivery, a problem that reflects the trend related to delivering groceries from the physical store (Delasay et al., 2022). Finally, recent research looked at the key role of delivery drivers in route-optimization problems. Liu et al. (2021) found that drivers' decision-making often deviates from the optimal route due to factors that the dispatching algorithm cannot observe, for example, traffic or weather conditions. In turn, this affects predictability of service time and uncertainty of last mile delivery, thus suggesting that algorithms should be integrated with operational data to consider the learning behavioral aspects of last mile delivery. Lu et al. (2020) and Lu, Suzuki, and Clottey (2022) studied the effective use of last mile delivery drivers' helpers and hybrid drivers to improve cost saving, driver time reduction, and delivery performance.

The literature has also explored innovations in last mile delivery by looking at implementing new delivery modes and integrating new delivery technologies to improve performance (Behnke, 2019; Na, Kweon, & Park, 2021; Dolan, 2022). Among new delivery technologies, delivery drones have received much attention in the literature. Perera et al. (2020)

analyzed the integration of delivery drones on an extant retailer's logistics, finding that increases of delivery speed is contingent upon the maturity of this technology usage. From this perspective, Maghazei, Lewis, and Netland (2022) conduct a longitudinal study to assess that drone delivery adoption is not a technology that follows a linear logic, instead it is governed by a dynamic interaction between the technological offer and the market demand to explore meaningful usage of the technology. Finally, Merkert, Bliemer, and Fayyaz (2022) identified consumers' preferences toward drone delivery as compared to traditional postal delivery, finding that customers positively evaluate the economic and operational advantages of drone delivery (e.g., cheaper and faster) but identify safety issues of the drop-off points. Another technology adopted in last mile delivery is the autonomous vehicles (AV). Reed et al. (2022) studied the integration of AV in last mile delivery operations, finding that AVs can reduce the total delivery time due to shorter operational times and increased capacity.

In addition to innovations in last mile delivery, prior literature also includes initiatives to improve performance based on the usage of delivery information to determine the optimal location and time window (Kull, Boyer, & Calantone, 2007). This reflects the need to improve performance and customer outcomes by leveraging data science to accurately predict customers' location and avoid failed deliveries (Praet & Martens, 2020), as well as a correct estimation of promised delivery time while maximizing retailer's outcomes. Other initiatives include improving environmentally sustainable deliveries by understanding customers' preferences for green time-window deliveries (Agatz, Fan, & Stam, 2021), consolidating drop-off points through parcel lockers (Wang, 2018; Lyu & Teo, 2022; Seghezzi et al., 2022), and eventually removing the last mile by encouraging consumer-pick-up at the store through self-collection (Hübner, Kuhn, &

Wollenburg, 2016; Li et al., 2018; Wollenburg et al., 2018; Wang et al., 2019a, 2019b; Lim et al., 2021).

Finally, the literature has looked at the delivery driver as a key resource. Specifically, this body of literature focused on driver familiarity with the delivery zone (in B2C) or a specific customer (in B2B). Driver familiarity increase service performance by reducing the service time (Zhong et al., 2007; Smilowitz et al., 2013). Indeed, the driver gains more experience in activities such as finding the address, parking, using shortcuts and alternative routes, and addressing the common issue of customers' missed deliveries (Schneider, 2016; Lyu & Teo, 2022; Seghezzi et al., 2022). Familiarity also reduces task complexity and reduces drivers' errors during delivery (Miller, Schwieterman, & Bolumole, 2018; Payyanadan, Sanchez, & Lee, 2019; Choi, 2020). Finally, driver familiarity also increases customer outcomes. Over multiple trips to the same customer, drivers establish a relationship with the customer and learn how to serve the customer best (Keller, 2002; Bode, Lindemann, & Wagner, 2011)

Overall, this literature does not fully explore the key role of driver familiarity in the crowdshipping context. Despite the delivery task being trivial, familiarity with the delivery zone offers improved service time, driver performance, and customer outcomes for on-demand delivery service. The experience of the crowdshipping driver is a competitive factor determining the success of the delivery (Mao et al., 2019, 2022). Hence, this study investigates how such learning experience influence delivery outcomes over time.

Theoretical background

We adopt logic from the body of knowledge on learning in service systems (Bimpikis & Markakis, 2019). This knowledge assumes that a service system processes tasks that may be heterogenous, utilizing servers with different skills that do not know the exact task type (Bimpikis & Markakis,

2019). Organizations optimally allocate resources in these service systems to match the task and the server's characteristics. For example, complex tasks are assigned to highly skilled healthcare professionals. Once the server is assigned to the task, he/she forms his/her beliefs about the task difficulty depending on whether the task was successfully completed. In addition, upon performing the task, he/she acquires knowledge and improves his/her learning relative to the task. Learning can assume the form of reducing the time required to complete a task while improving quality (Arlotto, Chick, & Gans, 2014). As servers are assigned to a multitude of tasks, they grow their experience and familiarity. This is supported by a worker's learning-by-doing, which suggests that as workers engage in multiple and different tasks, they gain experience that could be applied to other tasks (Staats & Gino, 2012). Indeed, being exposed to task variety initially increases the learning curve but eventually aids workers in applying the acquired knowledge to other tasks (Shafer et al., 2001; Staats & Gino, 2012). However, this logic does not account for the crowdshipping server's self-selection (Bimpikis & Markakis, 2019).

In crowdshipping, a driver engages in delivering an order based on a self-selection mechanism that allows the driver as an independent contractor to choose the task to perform. Thus, while the service system can provide service tasks, the crowdshipping driver ultimately decides whether or not to perform the delivery. This has important consequences for the delivery task. On the one hand, crowdshipping implies random service capacity, potentially decreasing the service quality (Benjaafar & Hu, 2020). On the other hand, delivery drivers can gain experience relative to a specific type of delivery task. In such cases, it is the driver, and not the platform (service system), to assign hard tasks to him/herself, without the mediation of the crowdshipping platform. Gaining experience over a task delivery results in increasing familiarity with the task (Gligor & Maloni, 2022). In the following, we develop our hypothesis focusing on the learning opportunity

related to driver's familiarity, and how familiarity impacts delivery outcomes. Figure 14 reports the theoretical model of this study.

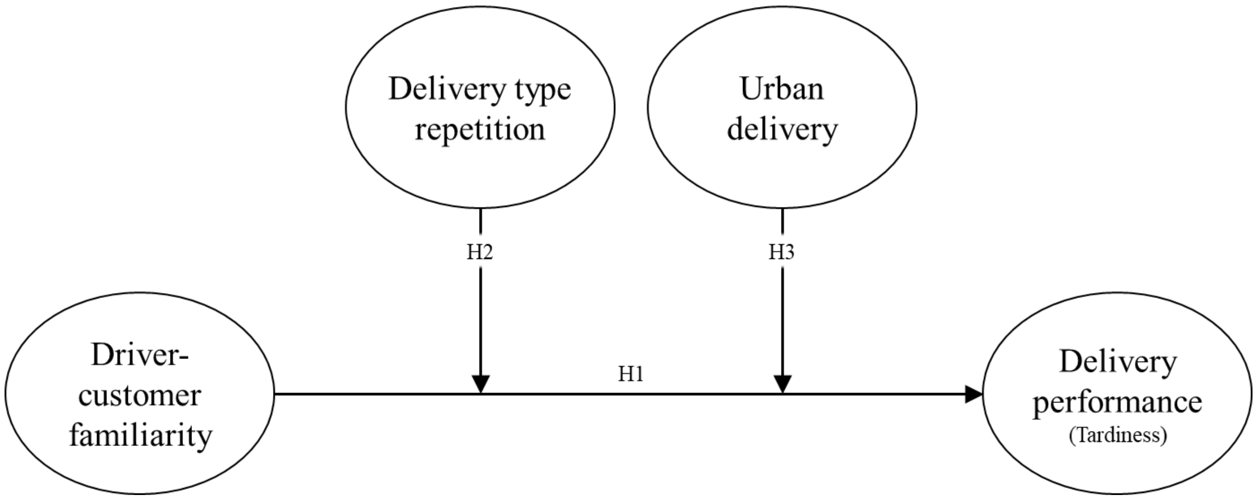


Figure 14 – Theoretical model essay 2

Hypotheses development

The impact of driver's familiarity on Delivery time performance

Delivery on-time performance refers to the timely delivery (Mentzer, Flint, & Kent, 1999; Akturk, Mallipeddi, & Jia, 2022). When a service provider perform a delivery task, he/she experiences several operational challenges that increase the service time, including finding the address, parking, and delivering to the desired destination (Liu et al., 2021). Once the delivery is completed, the driver creates experience and learns the characteristics and details of the specific delivery task. Over time, the driver performing the same delivery task acquires greater familiarity, which results in a greater level of comfort with the amount of service time needed for completion (Smilowitz et al., 2013), greater safety in the driving behaviors (e.g., no speeding) (Miller et al., 2018; Payyanadan et al., 2019), as well as faster in-site operations (Schneider, 2016). Hence, the more familiar is the driver with a delivery zone, the faster he/she can deliver to all customers in the delivery zone (Smilowitz et al., 2013). Thus, we expect that:

Hypothesis H1: *Driver's familiarity is positively associated with delivery time performance. A higher number of instances of repeating the delivery task to the same customer decrease tardiness.*

The moderating role of delivery type repetition

Cameron (2022) studies driver behaviors in service delivery, finding that crowd drivers either engage in a relational game in which they likely interact with customers to maximize the chances of tipping, or an efficiency game in which drivers maximize their utility by dedicating only to deliver packages, limiting the interaction with customers. Over time, drivers specialize in one of the two games and express preferences toward delivery tasks corresponding to a relational or efficiency game. Crowdshipping delivery tasks can either be attended or unattended (Lim & Srai, 2018). A driver performing either one of the two will become more familiar with the characteristics of the task. That is, repeated unattended delivery tasks will increase the chances of performing better when delivering, indeed, an unattended delivery. Conversely, a new delivery task of another type (attended) will decrease the performance of the driver. A similar logic holds for attended deliveries. Thus, a greater driver's fit to the task will likely moderate the direct effect of driver familiarity on delivery outcomes (Venkatraman, 1989). Hence, the driver will increase the delivery outcomes when the delivery is consistent with prior delivery characteristics. Thus, we expect that:

Hypothesis H2: *Delivery type repetition moderates the relationship between driver familiarity and delivery outcomes. That is, the higher number of repeated deliveries with the same characteristics, the stronger is the effect of driver familiarity on delivery time performance*.

Finally, the delivery area influences a driver's capability to acquire familiarity with the delivery task. A major difference between urban and rural deliveries is that in rural areas, drivers are often repeatedly assigned to the same delivery route, whereas in the urban context, delivery

drivers can be alternatively assigned to the core of the urban center or to the suburban areas (Rose et al., 2016, 2019, 2022). In addition, the delivery task in urban contexts may present higher uncertainty due to infrastructural complexities and higher levels of traffic and congestion (Rose et al., 2019). Uncertainty in the task decreases the chances of learning (Bimpikis & Markakis, 2019). Consequently, delivering in rural areas allows drivers to gain greater familiarity with a specific delivery task. In addition, drivers can decrease errors in rural areas as compared to urban contexts (Awaysheh et al., 2021). Finally, in urban contexts, the learning process may take longer given the difficulty of the task (Bimpikis & Markakis, 2019). Hence, a delivery task performed in urban contexts not only makes the learning process longer and more difficult, but errors and uncertainties may flaw the delivery performance. Thus, we expect that:

Hypothesis H3: *Delivery characteristics related to delivery area moderate the relationship between driver familiarity and delivery outcomes. That is, a delivery task performed in an urban context mitigates the positive effect of driver familiarity on delivery time performance*.

Empirical Setting and Data

Data description

We empirically investigated these hypotheses by compiling a dataset from multiple sources. First, we retrieved delivery operational data and driver outcomes data from a Fortune 100 retailer (hereafter called *Alpha*). The retailer has launched its white-label crowdshipping platform, which performs home deliveries from the retailer's stores using crowdsourced drivers. Upon being pinged from the retailer for home deliveries, the crowdshipping platform broadcasts the offer for the delivery task through the Driver App, where the offer is visible to a set of drivers assigned to a

driver zone[6]. The offer includes details related to the compensation, the pre-tipping, the delivery type, the number of orders included in the delivery task, the total of miles that a driver would drive to perform the task (i.e., from the store to the final drop-off), as well as the store and customers' address. Upon accepting the task, the Driver App provides GPS navigation instructions to arrive at the pick-up point and drop-off the orders at the customer's destination.

Alpha shared a raw dataset covering three months (February to April 2022), and containing approximately ~7 million customers' orders. The dataset includes detailed information about the characteristics of the order, customers, and drivers. The dataset reports several time stamps capturing the hour, minute, and second of the delivery operations for the order i, including order delivered ($orde_j$), which refers to the timestamp of when the driver delivered the order, and delivery time window end (de_j), which refers to the planned delivery time for a specific order. This dataset also includes the delivery type for each order (expedited, standard, regular), as well as the geo coordinates (latitude and longitude) of each drop-off location, as well as information relative to the store that dispatched the delivery task, including store id and complete address.

Second, we manually retrieved publicly available data from *Alpha*'s website about stores that dispatched a delivery task. Specifically, we downloaded data on stores' id, type, and exact addresses. *Alpha* sorts stores dispatching a delivery task into different[7] types, based on the dimension of the store. The exact address includes street name and number, 5-digit zip code, town, and state. Next, we built a custom program to extract the geographical coordinates (i.e., longitude and latitude) from each store address. Following current literature (Belo, Ferreira, & Telang, 2014; Belenzon, Chatterji, & Daley, 2020; Barrios, Hochberg, & Yi, 2022), we employed Google Maps

[6] As common in crowdshipping (Guda & Subramaniana, 2019; Tripathy, Bai, & Heese, 2022), the platform assigns drivers to a driver zone, which serves the stores located in at least one zip code (i.e., the same zone may serve multiple zip codes).
[7] For confidentiality reasons, the number of drivers, stores, zip codes, delivery zones, and store types are not disclosed.

Geocoding application programming interface (API), which is a Google Cloud Platform powered by Google Enterprise API, that converts between addresses and geographic coordinates. The platform library provides an XML file to automate the search for coordinates from the specified address and guarantees complete integration with Microsoft Excel. Among the advantages of this API is the correction of small variations in spelling (misspelling) between the manually retrieved and the actual address, allowing to find the geocoordinates when there is a close match (Belenzon et al., 2020). We manually checked a sample of these conversions to ensure the precision of the conversion, without finding any unreasonable matching (Belenzon et al., 2020). Finally, we matched this and *Alpha* datasets through store id. Hence, using the geocoordinates of drop-off locations (from *Alpha*) and of store locations (from Google), we computed the store-drop-off distance, expedited in miles, using *geodist* function in STATA17. This function computes ellipsoidal distances (i.e., "the length of the shortest curve between two points along the surface of the mathematical model of the earth WGS 1984 datum" (Picard, 2022) – the same used by Google Earth) using Vincenty (1975)'s equation (Picard, 2022) A limitation is that geodist computes the actual distance, not the travel distance. However, the nature of the dataset being limited to local store-to-home deliveries allows to reasonably assume that g_n and travel distance are similar and strongly correlated.

Third, we retrieved data relative to the categorization of Urban areas from the most recent classification of US Census Bureau based on the 2020 decennial census (US Census Bureau, 2022). The Census Bureau defines urban areas based on housing unit density measured at the census block level, with a minimum qualifying criteria of 2,000 housing units or a population of 5,000. The delineation process applies three housing unit densities: initial urban core of at least 435 housing units/m^2, remainder of urban area of at least 200 housing units/m^2, and at least one

high density nucleus of at least 1,275 housing units/m^2 (US Census Bureau, 2022). This classification has been used in prior literature investigating urban street networks (Boeing, 2020), ridehaling and traffic (Barrios et al., 2022).

Data cleaning

Before performing the main analyses, we cleaned the dataset following best practices of recent literature investigating similar contexts (Farber, 2015; Miao et al., 2022). First, we removed incomplete or erroneous observations and outliers for all the variables of interest. Specifically, we removed (1) orders with missing timestamps, (2) orders delivered by drivers whose age was below 21 and above 71[8], (3) erroneous observations with total travel distance longer than 21 miles[9] (4) erroneous observations of ordered delivered before 6 AM and after 10 PM[10]. In the end, 6,963,868 observations were retained as final sample. The unit of analysis of this study is at the single-order level.

Variables construction

We computed the outcome variables and predictors following theory, prior literature, and best practices. Table 13 reports the descriptive statistics, and Table 14 reports the correlations. The unit of analysis is at order level. This allows to determine the increasing driver familiarity throughout the dataset, as well as determine the impact of familiarity for the driver-customer dyad outcomes (tardiness for both, customer rating, driver retention).

---

[8] These thresholds result from symmetrically winsorizing driver age at 99%, align with prior literature investigating similar contexts (i.e., crowdshipping) (Ta, Esper, & Tokar, 2021), and are motivated by federal regulations on alcohol delivery, which may create unnecessary noise in the dataset (e.g., a 18 years old driver may not be given the choice of a delivery task of orders including alcohol delivery.

[9] This threshold result from symmetrically winsorized distance traveled at 99%, which produces an average distance traveled of 4.3 miles (SD = 4). This aligns with prior literature investigating similar context (i.e., crowdshipping), for example, (Miao et al., 2022) average trip distance was 11km (~7 miles), and (Castillo et al., 2022b)'s netnography report drivers performing between 7 and 10 miles.

[10] *Alpha* provides the delivery service from 7 AM to 9 PM.

Table 13 – Descriptive Statistics

	Variable	Mean	SD	Min	Max	Source
(1)	Tardiness (minutes)	7.17	21.86	0	125	(Atan et al., 2016)
(2)	Driver-Customer familiarity	1.13	0.49	1	45	(Smilowitz et al., 2013)
(3)	Delivery type repetition	0.04	0.2	0	1	(Ibanez et al., 2018)
(4)	Urban	0.54	0.5	0	1	
(5)	Customer distance from store (miles)*	3.48	2.25	0.29	9.64	(Akturk & Ketzenberg, 2022)
(6)	Customer subscription delivery service	0.69	0.46	0	1	(Wagner, 2021)
(7)	Order size	19.67	22.37	1	1212	(Liu et al., 2021)
(8)	Expedited	0.06	0.23	0	1	(Peinkofer, 2020)
(9)	Driver satisfaction	0.99	0.11	0	1	(Blaseg, 2020)
(10)	Driver fatigue	6.38	5.99	1	561	(Ergün-Şahin et al., 2022)
(11)	Driver restart	0.25	0.43	0	1	(Ibanez et al., 2018)
(12)	Driver age (years)*	42.31	11.05	21	71	(Ai et al., 2023)

Note: * The variable was symmetrically winsorized at 99%.

Table 14 – Correlations

	Variable	(1)	(2)	(3)	(4)	(5)	(6)	(7)
(1)	Tardiness (minutes)	1.00						
(2)	Driver-Customer familiarity	-0.04***	1.00					
(3)	Delivery type repetition	-0.04***	0.65***	1.00				
(4)	Urban	-0.01***	0.02***	0.02***	1.00			
(5)	Customer distance from store	0.27***	-0.05***	-0.05***	-0.07***	1.00		
(6)	Customer subscription delivery service	-0.16***	0.12***	0.11***	0.02***	-0.24***	1.00	
(7)	Order size	-0.16***	0.02***	0.03***	0.01***	-0.27***	0.23***	1.00
(8)	Expedited	-0.07***	0.01***	0.01***	0.02***	-0.12***	0.05***	0.03***
(9)	Driver satisfaction	-0.02***	0.01***	0.01***	0.00	-0.00	0.01***	-0.01***
(10)	Driver fatigue	0.07***	0.02***	0.01***	-0.00***	0.11***	-0.09***	-0.13***
(11)	Driver restart	-0.10***	-0.00***	0.00	0.00***	-0.19***	0.11***	0.13***
(12)	Driver age (years)*	-0.03***	0.04***	0.04***	-0.01***	-0.01***	0.02***	0.01***

	(8)	(9)	(10)	(11)	(12)
(8)	1.00				
(9)	-0.00***	1.00			
(10)	-0.06***	0.01***	1.00		
(11)	0.04***	-0.01***	-0.20***	1.00	
(12)	-0.00***	0.01***	0.04***	-0.03***	1.00

Note: * p < .1 ** p < .05 *** p < .01

The delivery performance reflects the on-time performance related to the delivery. We refer to the delivery performance using Tardiness. *Tardiness* is computed as a function of lateness and counts the number of minutes of delay an order has accumulated (Atan et al., 2016). For the order i, we operationalized Tardiness as follows;

$$Tardiness_i = \max(0, Lateness_i)$$

Where Lateness is the time elapsed between the actual delivery time stamp and the time stamp indicating the end of the delivery window (Thürer et al., 2015; Thürer, Fernandes, & Stevenson, 2020).

The independent variable, driver familiarity *(FAM)*, refers to the number of times a driver performed the delivery to a specific customer. The dataset allows matching the ID of the driver completing the order delivery with the ID of the customer receiving the delivery. Hence, following prior literature (Smilowitz et al., 2013), for the driver r delivering to the customer c the order i, FAM_{rci} was computed as follows:

$$FAM_{rci} = \sum_{r=1}^{R}\sum_{c=1}^{C}\sum_{i=1}^{I} n_{rci} - 1$$

Where n_{rci} denotes the ith order delivered by the driver *r* to the customer *c*. Greater values of FAM_{rci} express higher driver's familiarity delivering for the specific customer.

The moderator Delivery type repetition was captured using a dummy variable taking a value of 1 if the delivery type (unattended) performed to a customer was the same as the last delivery completed to the same customer, 0 otherwise. This operationalization follows prior literature (Ibanez et al., 2018), and reflects the learning experience of the delivery drivers, who can build delivery capabilities through learning.

The second moderator delivery characteristics in urban vs rural was captured using a dummy variable taking a value of 1 if the delivery area occurred in an urban context, 0 otherwise. We defined urban areas following (US Census Bureau, 2022), the most recent classification based on the 2020 decennial census.

We also include a set of control variables. First, we computed a set of control variables relative to the customer, namely the distance from the customer's residence to the store that

dispatched the order (Akturk & Ketzenberg, 2022), and whether the customer has subscribed to Alpha delivery service (Wagner et al., 2021). Next, we control for the order characteristics by including order size expressing the number of items in the order (Liu et al., 2021), and an indicator of an expedited delivery (Peinkofer et al., 2020). In addition, we included controls relative to the delivery drivers, such as age (Ai et al., 2023), driver satisfaction as a thumb-up measure (Blaseg et al., 2020), driver fatigue computed as the cumulative number of orders a driver delivers on a given day (Ergün-Şahin et al., 2022), and driver restart which determines is order i was the first in a day for the driver (Ibanez et al., 2018). Finally, we included fixed-effects for month, day of the week, clock hour of the day, and store type.

Results

Preliminary analysis

Prior to the analysis, we carefully review the distribution of the dependent variable and prior literature to select the best estimation model for each dv. In addition, following best practices (Aiken, West, & Reno, 1991) and common procedures in the operations management literature (e.g., Kim & Zhu, 2018; Amengual & Apfelbaum, 2021; Delfgaauw et al., 2022), in estimating the effect of interactions between continuous independent variables, prior to the empirical analysis, we mean-centered the focal predictors.

Estimation Models and Results

To test the hypothesis with Tardiness as dependent variable, we adopted a Poisson regression model because Tardiness presents a Poisson distribution. H1 predicted a positive relationship between Driver-Customer Familiarity and Delivery Time Performance, hence a negative relationship with Tardiness. We tested this hypothesis as follows:

$$(17)\ \Pr(Tardiness_i = s|X_i) = Poisson(\alpha + \beta FAM_{rci} + \mathrm{B}\Gamma_i + \varepsilon_i)$$

Where Γ_i is the vector of independent variables, B is the vector of coefficients, and ε_i is the error term clustered on driver ID to reduce potential source of heteroscedasticity. Γ_iB is specified as follows:

$$(18)\ \Gamma_i = \gamma_0 + \gamma_1 DistStore_c + \gamma_2 Age_d + \gamma_3 Sub_c + \gamma_4 DistStore_c + \gamma_5 Size_i + \gamma_6 Exp_i + \gamma_8 Fatigue_d + \gamma_{9is} store_{is} + \gamma_{10im} month_{im} + \gamma_{11iw} dow_{iw} + \gamma_{14ih} hh_{ih}$$

Results in Table 15 (Model 1) show a negative significant coefficient of Familiarity on Tardiness (β = -0.150), supporting H1. For an additional unit of familiarity (i.e., an additional order that the driver delivers to the same customer), Tardiness reduces by approximately 14%. Next, we tested H3a and H4a relative to the moderation effects of Delivery type repetition and Urban, by utilizing the same Poisson regression estimation model but including the two moderators and the interaction terms into the equation. Results in Table 15 (Model 2) show a negative direct effect of Delivery type repetition but a positive effect of Urban on Tardiness, and the interaction terms in Model 3 suggest differential moderation effects. Specifically, H3 predicted that a driver repeating the same task would enhances the positive effect of familiarity on delivery time performance. We found that when drivers repeat the same type of delivery for a customer, tardiness decreases, but as a driver become more and more familiar with the customer, the differential effect between repeating the same delivery task and performing an alternative task type decreases. Figure 15 suggests a 1.5 minutes of difference between a repeated vs unrepeated task. H4 predicted that deliveries in Urban context were more challenging for delivery drivers, hence mitigating the effect of familiarity on tardiness. Figure 16 suggests that at low levels of familiarity, delivering in the urban context presents higher levels of tardiness, but as a driver becomes more familiar with delivering to a customer, the delivery area does not seem to play a role (i.e., the two curves seem to converge). Hence, we find support for H4.

Table 15 – Poisson Regression: The effect of Familiarity, Delivery type repetition, Repeat urban, and Interactions on Tardiness

	(1) Tardiness β	se	(2) Tardiness β	se	(3) Tardiness β	se
Distance from store	0.215****	(0.001)	0.215****	(0.001)	0.215****	(0.001)
Driver age	-0.001****	(0.000)	-0.001****	(0.000)	-0.001****	(0.000)
Customer subscription	-0.383****	(0.003)	-0.382****	(0.003)	-0.382****	(0.003)
Order size	-0.023****	(0.000)	-0.023****	(0.000)	-0.023****	(0.000)
Expedited	-1.760****	(0.018)	-1.761****	(0.018)	-1.761****	(0.018)
Driver fatigue	0.005****	(0.001)	0.005****	(0.001)	0.005****	(0.001)
Store fe	YES		YES		YES	
Month fe	YES		YES		YES	
Delivery day of the week fe	YES		YES		YES	
Delivery hour fe	YES		YES		YES	
Familiarity	-0.150****	(0.005)	-0.110****	(0.005)	-0.118****	(0.008)
Delivery type repetition			-0.158****	(0.010)	-0.218****	(0.015)
Urban			0.043****	(0.005)	0.042****	(0.005)
Repeat unattended x Familiarity					0.064****	(0.011)
Urban x Familiarity					-0.019*	(0.009)
Constant	3.440****	(0.126)	3.417****	(0.126)	3.415****	(0.126)
χ^2	336473.083		338124.767		338289.853	
Pseudo-R2	0.297		0.298		0.297	
N	6,963,868		6,963,868		6,963,868	

Note: $^{+}$ $p < 0.1$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.005$, **** $p < 0.001$; Reported robust standard errors are clustered on driver ID.

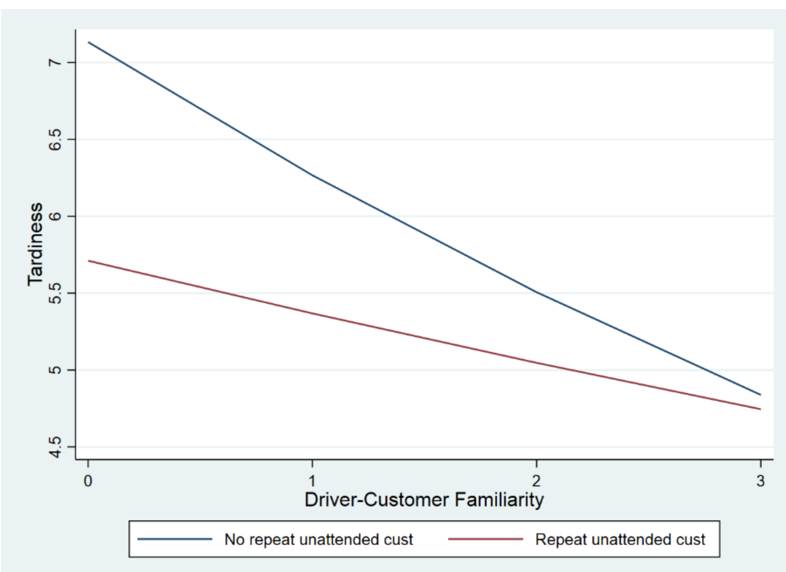


Figure 15 – Interaction between Driver-Customer Familiarity and Repeat Delivery Type (Table 15 Model 3)

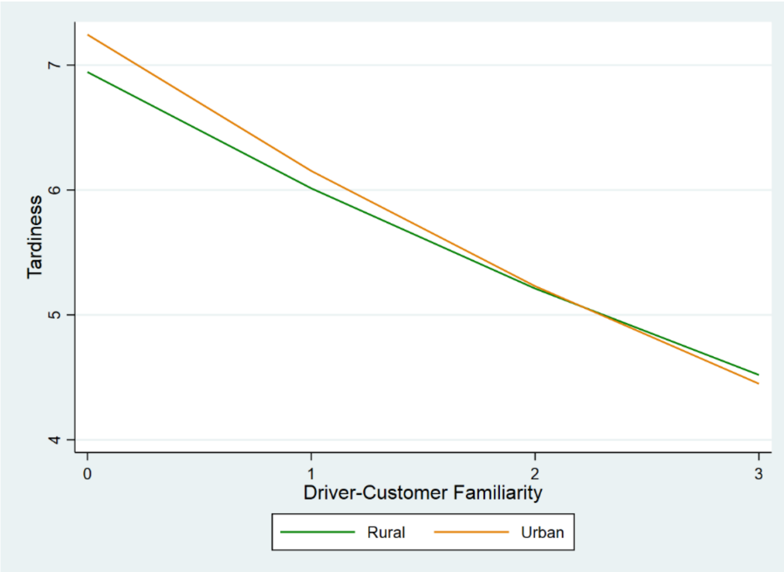


Figure 16 – Interaction between Driver-Customer Familiarity and Urban (Table 15 Model 3)

To assess the robustness of these estimations, we performed a series of robustness checks (Appendix B). First, we addressed endogeneity relative to the estimation model, by performing the same analysis but replacing the dependent variable Tardiness with a binary variable Delivery Performance, taking 1 if the order was late, 0 otherwise, and adopting a Logit Regression. Estimations reported in Table 38 show consistent results. Next, we performed the same analysis of Table 15, but including an additional control variable (Restart), which is an indicator taking 1 if order i was the first in a given day for the driver. Prior literature discusses the importance of the *warm-up* effect that a worker may suffer on the first task of the day (Ibanez et al., 2018). Estimations included in Table 39 report similar results.

Discussion

This set of results suggests that familiarity plays an important role in crowdshipping deliveries time performance, but the positive effect of familiarity on delivery time performance is affected by unattended deliveries and urban context. Results reports that the learning effect relative to the delivery type is contingent upon delivering to a specific customer. Indeed, when a driver performs

the same delivery type toward the same customer, deliveries present a better delivery time performance. This is especially important for unattended deliveries, which are inherently more difficult for delivery drivers, who occasionally fail unattended deliveries due to the greater complexity of finding reception boxes or customers' address (Lim, Wang, & Webster, 2023; Olsson, Hellström, & Vakulenko, 2023). Conversely, urban deliveries present drivers with challenges that only familiarity seems to address. Hence, traffic, congestion, limited parking, and more chances for human errors (Österle et al., 2015; Rose et al., 2019; Dayarian & Savelsbergh, 2020; Awaysheh et al., 2021), decrease the time performance especially at low levels of driver-customer familiarity.

These results contribute to prior literature and theory in many ways. First, we contribute to the literature on driver's familiarity by determining its impact in the crowdshipping context, and address recent calls for a greater understanding of how driver's knowledge of a service area or of customers' needs improve delivery performance (Mao et al., 2022). While prior literature investigated how crowdshipping driver familiarity influence customer purchasing behaviors (Mao et al., 2019), to the best of our knowledge this is the first study that matches crowdsourced driver's familiarity with a customer, and determine the operational performance for the single delivery. This is important because we provide additional evidence that expands the scarce literature in crowdshipping deliveries. Next, we also contribute to the body of literature studying crowdshipping driver's preferences relative to a delivery type. Cameron (2022) suggests that drivers improve their performance as they develop a preference for a specific delivery type. We found that the learning experience is valuable for a driver and improves performance only when the driver becomes more familiar with the customer and with the delivery type. Finally, we also contribute to the rising literature discussing delivery performance in urban deliveries (Rose et al.,

2019; Dayarian & Savelsbergh, 2020; Castillo et al., 2022a), and address recent calls for research focusing on how crowdshipping models work in urban vs rural areas (Ta et al., 2023). We expand this literature by comparing the performance of crowdshipping in rural vs ruban areas. Ermagun, Shamshiripour, & Stathopoulos (2020) found that crowdshipping in urban areas present greater performance because urban areas, with greater density and chances of vehicle ownership, ensure drivers capacity. In contrast, we found that delivering in urban areas present lower time performance.

We also offer additional evidence to inform the theoretical gap in the literature investigating workers' task familiarity (Shafer et al., 2001; Staats & Gino, 2012; Bimpikis & Markakis, 2019). On the one hand, exposing servers repeatedly to the same task decreases the learning curve, increases delivery performance but worsens server's performance on other tasks (Haughton, 2008; Mao et al., 2022). On the other hand, exposing servers to task variety increases the learning curve but servers can use the acquired knowledge on other tasks (Zhong et al., 2007). Our results align with the former. Familiarity with a customer decreases the learning curve and improve performance, and, an exposure to task variety does not seem to improve the delivery performance.

Finally, managerially, we provide evidence that matching drivers to customers improve the platform's operational performance. Specifically, we found that for an additional delivery that a driver performs to a specific customer, lateness reduces by 14%. In a context where the driver selects the task without prior knowledge of whom the order is to be delivered, we recommend the platform to consider matching drivers to the same customer or a set of customers. Familiarity can also address the challenges related to unattended deliveries and urban logistics. While unattended deliveries present benefits related to longer delivery windows, a more efficient delivery planning,

and lower failed delivery rate (Olsson et al., 2023), the actual service provision present difficulties for the driver related to missing customers' presence. First, during attended deliveries, drivers typically know that customers are waiting their order at the final destination. Hence, drivers may be motivated to improve their on time performance. Second, the dropping-off, especially in urban areas, may require additional time and extra effort of a driver when the delivery is unattended. For example, a driver may find difficulties related to accessing a building and finding the correct front door due to the building policies (Kaysen, 2023). To address these issues, Amazon offered customers incentives to adopt Key devices that grant drivers access to the lobby (Hamilton, 2021).

Limitations and Future Research

As for any empirical research, this study also presents limitations. First, we recognize the limitations of using a dataset from a single industry partner. Although this allows to remove the variance affected by idiosyncrasies across companies, hence cleaner results (Lu, Lee, & Son, 2022a), and we complemented the dataset with publicly available data, the truncated dataset does not allow to observe the whole lifespan of drivers and customers outside the time period. We encourage researchers and practitioners to cautiously interpret the conclusions of this study. Second, we acknowledge that monetary incentives may be affecting the delivery driver performance and driver retention (Castillo et al., 2022b). While we control for the number of deliveries completed by a driver on a given day, elements such as customer's tip or overall remuneration may be factors determining, besides familiarity, improvements in the dependent variables. Thus, we encourage future research to explore the combined effect of driver's familiarity and monetary incentives on, for example, determining the delivery time performance of a driver for attended deliveries. Finally, we acknowledge the limitations of some variables operationalizations, such as the binary variable Urban deliveries. While we identified urban areas

adopting the U.S. Bureau official list of urban areas, crowdshipping platforms may follow other categorizations and share urban areas in multiple delivery zones. Future research could investigate how familiarity with delivery zones, instead of familiarity with a customer, could improve the performance of the driver.

Reference list

Agatz, N., Fan, Y., & Stam, D. 2021. The Impact of Green Labels on Time Slot Choice and Operational Sustainability. *Production and Operations Management*, 30(7): 2285–2303.

Ai, W., Chen, Y., Mei, Q., Ye, J., & Zhang, L. 2023. Putting Teams into the Gig Economy: A Field Experiment at a Ride-Sharing Platform. *Management Science*, (February): 1–43.

Aiken, L. S., West, S. G., & Reno, R. R. 1991. *Multiple regression: Testing and interpreting interactions*. London: Sage.

Akturk, M. S., & Ketzenberg, M. 2022. Exploring the Competitive Dimension of Omnichannel Retailing. *Management Science*, 68(4): 2732–2750.

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

Amengual, M., & Apfelbaum, E. P. 2021. True motives: Prosocial and instrumental justifications for behavioral change in organizations. *Management Science*, 67(8): 5032–5051.

Arlotto, A., Chick, S. E., & Gans, N. 2014. Optimal Hiring and Retention Policies for Heterogeneous Workers Who Learn. *Management Science*, 60(1): 110–129.

Atan, Z., De Kok, T., Dellaert, N. P., Van Boxel, R., & Janssen, F. 2016. Setting planned leadtimes in customer-order-driven assembly systems. *Manufacturing and Service Operations Management*, 18(1): 122–140.

Awaysheh, A., Frohlich, M. T., Flynn, B. B., & Flynn, P. J. 2021. To err is human: Exploratory multilevel analysis of supply chain delivery delays. *Journal of Operations Management*, 67(7): 882–916.

Barrios, J. M., Hochberg, Y. V., & Yi, H. 2022. The cost of convenience: Ridehailing and traffic fatalities. *Journal of Operations Management*, (September 2021): 1–33.

Behnke, M. 2019. Recent Trends in Last Mile Delivery: Impacts of Fast Fulfillment, Parcel Lockers, Electric or Autonomous Vehicles, and More. In C. Bierwirth, T. Kirschstein, & D. Sackman (Eds.), *Logistics Management* (pp. 141–156). Springer Berlin Heidelberg.

Belenzon, S., Chatterji, A. K., & Daley, B. 2020. Choosing between growth and glory. *Management Science*, 66(5): 2050–2074.

Belo, R., Ferreira, P., & Telang, R. 2014. Broadband in school: Impact on student performance. *Management Science*, 60(2): 265–282.

Benjaafar, S., & Hu, M. 2020. Operations management in the age of the sharing economy: What is old and what is new? *Manufacturing and Service Operations Management*, 22(1): 93–101.

Bimpikis, K., & Markakis, M. G. 2019. Learning and hierarchies in service systems. *Management Science*, 65(3): 1268–1285.

Blaseg, D., Schulze, C., & Skiera, B. 2020. Consumer protection on Kickstarter. *Marketing Science*, 39(1): 211–233.

Bode, C., Lindemann, E., & Wagner, S. M. 2011. Driving trucks and driving sales? the impact of delivery personnel on customer purchase behavior. *Journal of Business Logistics*, 32(1): 99–114.

Boeing, G. 2020. A multi-scale analysis of 27,000 urban street networks: Every US city, town, urbanized area, and Zillow neighborhood. *Environment and Planning B: Urban Analytics and City Science*, 47(4): 590–608.

Cachon, G. P., Daniels, K. M., & Lobel, R. 2017. The Role of Surge Pricing on a Service Platform with Self-Scheduling Capacity. *Manufacturing & Service Operations Management*, 19(3): 368–384.

Cameron, L. D. 2022. “Making Out” While Driving: Relational and Efficiency Games in the Gig Economy. *Organization Science*, 33(1): 231–252.

Caro, F., Kök, A. G., & Martínez-de-Albéniz, V. 2020. The Future of Retail Operations. *Manufacturing & Service Operations Management*, 22(1): 47–58.

Castillo, V. E., Bell, J. E., Mollenkopf, D. A., & Stank, T. P. 2022a. Hybrid last mile delivery fleets with crowdsourcing: A systems view of managing the cost-service trade-off. *Journal of Business Logistics*, 43(1): 36–61.

Castillo, V. E., Mollenkopf, D. A., Bell, J. E., & Esper, T. L. 2022b. Designing technology for on-demand delivery: The effect of customer tipping on crowdsourced driver behavior and last mile performance. *Journal of Operations Management*, 68(5): 424–453.

Choi, J. C. 2020. *User Familiarity and Satisfaction With Food Delivery Mobile Apps* (Vol. 10).

Convey. 2018. *Last Mile Delivery: What Shoppers Want and How to #SaveRetail*. *Convey*. Retrieved from https://www.getconvey.com/press-d-last-mile-delivery-save-retail/

Cui, R., Li, M., & Li, Q. 2020. Value of high-quality logistics: Evidence from a clash between SF express and alibaba. *Management Science*, 66(9): 3879–3902.

Cui, V., Yang, H., & Vertinsky, I. 2018. Attacking your partners: Strategic alliances and competition between partners in product markets. *Strategic Management Journal*, 39(12): 3116–3139.

Dayarian, I., & Savelsbergh, M. 2020. Crowdshipping and Same-day Delivery: Employing In-store Customers to Deliver Online Orders. *Production and Operations Management*, 29(9): 2153–2174.

Delasay, M., Jain, A., & Kumar, S. 2022. Impacts of the COVID-19 pandemic on grocery retail operations: An analytical model. *Production and Operations Management*, (September

2020).

Delfgaauw, J., Dur, R., Onemu, O., & Sol, J. 2022. Team Incentives, Social Cohesion, and Performance: A Natural Field Experiment. *Management Science*, 68(1): 230–256.

Dolan, S. 2022. The challenges of last mile delivery logistics and the tech solutions cutting costs in the final mile. *Business Insider*. Retrieved November 1, 2022, from https://www.businessinsider.com/last-mile-delivery-shipping-explained

Ergün-Şahin, B., Güneş, E. D., Kocabıyıkoğlu, A., & Keskin, A. 2022. How does workload affect test ordering behavior of physicians? An empirical investigation. *Production and Operations Management*, (Forthcoming).

Ermagun, A., Shamshiripour, A., & Stathopoulos, A. 2020. *Performance analysis of crowd-shipping in urban and suburban areas*. *Transportation* (Vol. 47). Springer US.

Farber, H. S. 2015. Why you can't find a taxi in the rain and other labor supply lessons from cab drivers. *Quarterly Journal of Economics*, 130(4): 1975–2026.

Fatehi, S., & Wagner, M. R. 2022. Crowdsourcing Last-Mile Deliveries. *Manufacturing & Service Operations Management*, 24(2): 791–809.

Gligor, D. M., & Maloni, M. J. 2022. More is not always better: The impact of value co-creation fit on B2B and B2C customer satisfaction. *Journal of Business Logistics*, 43(2): 209–237.

Guda, H., & Subramaniana, U. 2019. Your uber is arriving: Managing on-demand workers through surge pricing, forecast communication, and worker incentives. *Management Science*, 65(5): 1995–2014.

Hamilton, I. A. 2021. Amazon is offering building managers $100 gift cards as an incentive to install "Key" devices, which let delivery drivers get into the lobby at all hours. *Business Insider*. Retrieved October 4, 2023, from https://www.businessinsider.com/amazon-key-for-business-delivery-drivers-packages-building-managers-2021-7

Haughton, M. A. 2008. The efficacy of exclusive territory assignments to delivery vehicle drivers. *European Journal of Operational Research*, 184(1): 24–38.

Hsiao, Y. H., Chen, M. C., Lu, K. Y., & Chin, C. L. 2018. Last-mile distribution planning for fruit-and-vegetable cold chains. *International Journal of Logistics Management*, 29(3): 862–886.

Hübner, A., Kuhn, H., & Wollenburg, J. 2016. Last mile fulfilment and distribution in omni-channel grocery retailing: A strategic planning framework. *International Journal of Retail and Distribution Management*, 44(3): 228–247.

Ibanez, M. R., Clark, J. R., Huckman, R. S., & Staats, B. R. 2018. Discretionary task ordering: Queue management in radiological services. *Management Science*, 64(9): 4389–4407.

Kaysen, R. 2023. I'm Ready for Deliveries at My Door Again, but My Building Still Says No. *The New York Times*. Retrieved October 4, 2023, from https://www.nytimes.com/2023/03/11/realestate/food-delivery-apartment-buildings.html

Keller, S. B. 2002. Driver Relationships With Customers and Driver Turnover: Key Mediating Variables Affecting Driver Performance in the Field. *Journal of Business Logistics*, 23(1):

39–64.

Kim, D. Y., & Zhu, P. 2018. Supplier dependence and R&D intensity: The moderating role of network centrality and interconnectedness. *Journal of Operations Management*, 64(September 2017): 7–18.

Kull, T. J., Boyer, K., & Calantone, R. 2007. Last-mile supply chain efficiency: An analysis of learning curves in online ordering. *International Journal of Operations and Production Management*, 27(4): 409–434.

Li, Y., Liu, H., Lim, E. T. K., Goh, J. M., Yang, F., & Lee, M. K. O. 2018. Customer's reaction to cross-channel integration in omnichannel retailing: The mediating roles of retailer uncertainty, identity attractiveness, and switching costs. *Decision Support Systems*, 109: 50–60.

Lim, S. F. W. T., Jin, X., & Srai, J. S. 2018. Consumer-driven e-commerce: A literature review, design framework, and research agenda on last-mile logistics models. *International Journal of Physical Distribution and Logistics Management*, 48(3): 308–332.

Lim, S. F. W. T., Rabinovich, E., Park, S., & Hwang, M. 2021. Shopping Activity at Warehouse Club Stores and Its Competitive and Network Density Implications. *Production and Operations Management*, 30(1): 28–46.

Lim, S. F. W. T., & Srai, J. S. 2018. Examining the anatomy of last-mile distribution in e-commerce omnichannel retailing: A supply network configuration approach. *International Journal of Operations and Production Management*, 38(9): 1735–1764.

Lim, S. F. W. T., Wang, Q., & Webster, S. 2023. Do it right the first time: Vehicle routing with home delivery attempt predictors. *Production and Operations Management*, Forthcomin.

Liu, S., He, L., & Shen, Z. J. M. 2021. On-time last-mile delivery: Order assignment with travel-time predictors. *Management Science*, 67(7): 4095–4119.

Lu, G., Lee, H. S., & Son, J. 2022a. Product variety in local grocery stores: Differential effects on stock-keeping unit level sales. *Journal of Operations Management*, 68(1): 33–54.

Lu, S. H., Suzuki, Y., & Clottey, T. 2020. The Last Mile: Managing Driver Helper Dispatching for Package Delivery Services. *Journal of Business Logistics*, 41(3): 206–221.

Lu, S. H., Suzuki, Y., & Clottey, T. 2022b. Improving the efficiency of last-mile package deliveries using hybrid driver helpers. *Decision Sciences*, (September 2021): 1–23.

Lyu, G., & Teo, C.-P. 2022. Last Mile Innovation: The Case of the Locker Alliance Network. *Manufacturing & Service Operations Management*, (Forthcoming).

Maghazei, O., Lewis, M. A., & Netland, T. H. 2022. Emerging technologies and the use case: A multi-year study of drone adoption. *Journal of Operations Management*, (Forthcoming).

Mao, W., Ming, L., Rong, Y., Tang, C. S., & Zheng, H. 2019. Faster Deliveries and Smarter Order Assignments for an On-Demand Meal Delivery Platform. *SSRN Electronic Journal*.

Mao, W., Ming, L., Rong, Y., Tang, C. S., & Zheng, H. 2022. On-Demand Meal Delivery Platforms: Operational Level Data and Research Opportunities. *Manufacturing & Service*

Operations Management, (May).

McKinnon, A. C., & Tallam, D. 2003. Unattended delivery to the home: An assessment of the security implications. *International Journal of Retail & Distribution Management*, 31(1): 30–41.

Mentzer, J. T., Flint, D. J., & Kent, J. L. 1999. Developing a logistics service quality scale. *Journal of Business Logistics*, 9(2): 1–47.

Merkert, R., Bliemer, M. C. J., & Fayyaz, M. 2022. Consumer preferences for innovative and traditional last-mile parcel delivery. *International Journal of Physical Distribution & Logistics Management*, 52(3): 261–284.

Miao, W., Deng, Y., Wang, W., Liu, Y., & Tang, C. S. 2022. The effects of surge pricing on driver behavior in the ride-sharing market: Evidence from a quasi-experiment. *Journal of Operations Management*, Forthcomin.

Miller, J. W., Schwieterman, M. A., & Bolumole, Y. A. 2018. Effects of Motor Carriers' Growth or Contraction on Safety: A Multiyear Panel Analysis. *Journal of Business Logistics*, 39(2): 138–156.

Na, H. S., Kweon, S. J., & Park, K. 2021. Characterization and Design for Last Mile Logistics: A Review of the State of the Art and Future Directions. *Applied Sciences*, 12(1): 118.

Olsson, J., Hellström, D., & Vakulenko, Y. 2023. Customer experience dimensions in last-mile delivery: an empirical study on unattended home delivery. *International Journal of Physical Distribution & Logistics Management*, 53(2): 184–205.

Österle, I., Aditjandra, P. T., Vaghi, C., Grea, G., & Zunder, T. H. 2015. The role of a structured stakeholder consultation process within the establishment of a sustainable urban supply chain. *Supply Chain Management: An International Journal*, 20(3): 284–299.

Paradiso, R., Roberti, R., Lagana, D., & Dullaert, W. 2020. An exact solution framework for multitrip vehicle-routing problems with time windows. *Operations Research*, 68(1): 180–198.

Payyanadan, R. P., Sanchez, F. A., & Lee, J. D. 2019. Influence of familiarity on the driving behavior, route risk, and route choice of older drivers. *IEEE Transactions on Human-Machine Systems*, 49(1): 10–19.

Peinkofer, S. T., Schwieterman, M. A., & Miller, J. W. 2020. Last-mile delivery in the motor-carrier industry: A panel data investigation using discrete time event history analysis. *Transportation Journal*, 59(3): 129–164.

Perera, S., Dawande, M., Janakiraman, G., & Mookerjee, V. 2020. Retail Deliveries by Drones: How Will Logistics Networks Change? *Production and Operations Management*, 29(9): 2019–2034.

Picard, R. 2022. help geodist. Retrieved from http://fmwww.bc.edu/repec/bocode/g/geodist.html

Praet, S., & Martens, D. 2020. Efficient Parcel Delivery by Predicting Customers' Locations. *Decision Sciences*, 51(5): 1202–1231.

Punakivi, M., Yrjölä, H., & Holmström, J. 2001. Solving the last mile issue: Reception box or delivery box? *International Journal of Physical Distribution and Logistics Management*, 31(6): 427–439.

Reed, S., Campbell, A. M., & Thomas, B. W. 2022. The Value of Autonomous Vehicles for Last-Mile Deliveries in Urban Environments. *Management Science*, 68(1): 280–299.

Rose, W. J., Bell, J. E., & Griffis, S. E. 2022. Inductive research in last-mile delivery routing: Introducing the Re-Gifting heuristic. *Journal of Business Logistics*, (Forthcoming).

Rose, W. J., Mollenkopf, D. A., Autry, C. W., & Bell, J. E. 2016. Exploring urban institutional pressures on logistics service providers. *International Journal of Physical Distribution & Logistics Management*, 46(2): 153–176.

Rose, W. J., Mollenkopf, D. A., Autry, C. W., & Williams, B. D. 2019. Tailoring Transportation Planning Decisions to Diverse Urban Environments. *Decision Sciences*, 51(4): 920–961.

Schneider, M. 2016. The vehicle-routing problem with time windows and driver-specific times. *European Journal of Operational Research*, 250(1): 101–119.

Seghezzi, A., Siragusa, C., & Mangiaracina, R. 2022. Parcel lockers vs. home delivery: a model to compare last-mile delivery cost in urban and rural areas. *International Journal of Physical Distribution and Logistics Management*, 52(3): 213–237.

Shafer, S. M., Nembhard, D. A., & Uzumeri, M. V. 2001. The effects of worker learning, forgetting, and heterogeneity on assembly line productivity. *Management Science*, 47(12): 1639–1653.

Shang, W., & Liu, L. 2011. Promised delivery time and capacity games in time-based competition. *Management Science*, 57(3): 599–610.

Smilowitz, K., Nowak, M., & Jiang, T. 2013. Workforce management in periodic delivery operations. *Transportation Science*, 47(2): 214–230.

Staats, B. R., & Gino, F. 2012. Specialization and variety in repetitive tasks: Evidence from a Japanese bank. *Management Science*, 58(6): 1141–1159.

Stroh, A. M., Erera, A. L., & Toriello, A. 2022. Tactical Design of Same-Day Delivery Systems. *Management Science*, 68(5): 3444–3463.

Ta, H., Esper, T. L., Rossiter Hofer, A., & Sodero, A. 2023. Crowdsourced delivery and customer assessments of e-Logistics Service Quality: An appraisal theory perspective. *Journal of Business Logistics*, (Forthcoming).

Ta, H., Esper, T. L., & Tokar, T. 2021. Appealing to the Crowd: Motivation Message Framing and Crowdsourcing Performance in Retail Operations. *Production and Operations Management*, 30(9): 3192–3212.

Thürer, M., Fernandes, N. O., & Stevenson, M. 2020. Material Flow Control in High-Variety Make-to-Order Shops: Combining COBACABANA and POLCA. *Production and Operations Management*, 29(9): 2138–2152.

Thürer, M., Land, M. J., Stevenson, M., Fredendall, L. D., & Godinho Filho, M. 2015. Concerning

Workload Control and Order Release: The Pre-Shop Pool Sequencing Decision. *Production and Operations Management*, 24(7): 1179–1192.

Tian, X., Shi, J., & Qi, X. 2022. Stochastic Sequential Allocations for Creative Crowdsourcing. *Production and Operations Management*, 31(2): 697–714.

Tripathy, M., Bai, J., & Heese, H. S. 2022. Driver collusion in ride-hailing platforms. *Decision Sciences*, (March 2021): 1–13.

US Census Bureau Census Bureau Notices - Final Criteria: Urban Area Criteria for the 2020 Census. , 87 Federal Register 16706–16715 (2022).

Venkatraman, N. 1989. The Concept of Fit in Strategy Research: Toward Verbal and Statistical Correspondence. *Academy of Management Review*, 14(3): 423–444.

Vincenty, T. 1975. Direct and Inverse Solutions of Geodesics on the Ellipsoid with Application of Nested Equations. *Survey Review*, 23(176): 88–93.

Wagner, L., Pinto, C., & Amorim, P. 2021. On the Value of Subscription Models for Online Grocery Retail. *European Journal of Operational Research*, 294(3): 874–894.

Wang, X. 2018. An innovation diffusion perspective of e-consumers ’ initial adoption of self-collection service via automated parcel station. *The International Journal of Logistics Management*, 29(1): 237–260.

Wang, X., Wong, Y. D., Teo, C. C., Yuen, K. F., & Li, K. X. 2019a. Decomposing service conveniences in self-collection: An integrated application of the SERVCON and Kano models. *International Journal of Physical Distribution and Logistics Management*, 49(4): 356–375.

Wang, X., Yuen, K. F., Wong, Y. D., & Teo, C. C. 2019b. Consumer participation in last-mile logistics service: an investigation on cognitions and affects. *International Journal of Physical Distribution and Logistics Management*, 49(2): 217–238.

Wollenburg, J., Hübner, A., Kuhn, H., & Trautrims, A. 2018. From bricks-and-mortar to bricks-and-clicks: Logistics networks in omni-channel grocery retailing. *International Journal of Physical Distribution and Logistics Management*, 48(4): 415–438.

Zhong, H., Hall, R. W., & Dessouky, M. 2007. Territory planning and vehicle dispatching with driver learning. *Transportation Science*, 41(1): 74–89.

IV. Essay 3: Addressing consumers' preferences in last mile delivery: The impact of delivery time performance and order characteristics on customer outcomes

Introduction

Meeting customers' expectations on last mile delivery service has become important to improve customer outcomes (Daugherty et al., 2019). Customers expect retail supply chains to provide this service at their convenience, with timely, quick, and punctual deliveries (Gawor & Hoberg, 2019; Mangiaracina et al., 2019). A timely – on-time – delivery is a critical component of the online shopping experience, because customers assess the value of purchasing from a retailer (Rao, Goldsby, et al., 2011; Wan et al., 2016; Akturk et al., 2022). Delivery speed – the interval between a customer's order and the delivery, is an essential service component of customer satisfaction (M. Fisher et al., 2015). Hence, retailers offer several delivery speed options to win online customer orders, such as expedited delivery, same-day delivery, and next-day delivery (Ishfaq et al., 2016; Peinkofer et al., 2020; Stroh et al., 2022). Finally, delivery punctuality, defined as receiving the package within an established delivery time lapse (Mangiaracina et al., 2019), is highly valued by consumers, yet it becomes problematic due to customized delivery windows, which represents a constraint when planning last mile deliveries (Xing et al., 2010; Lim & Winkenbach, 2019; Nguyen et al., 2019; Praet & Martens, 2020). Thus, providing a delivery service that meets customers' expectations has become increasingly relevant (Zimmermann et al., 2020). However, despite its importance, research has yet to unveil the impact of last mile delivery performance on actual customers' behaviors (Ketchen et al., 2021).

A growing body of supply chain and operations management literature has investigated customers' expectations on the order fulfillment process and last mile delivery. A stream of literature investigated how customers' expectations relative to the physical distribution service

quality shape customers outcomes, for example, satisfaction and repurchase intentions (Rabinovich et al., 2008; Davis-Sramek et al., 2010; Rao, Griffis, et al., 2011; Rao et al., 2014). Related research built on this knowledge by investigating how the forming of expectations toward the physical distribution service quality depends on the order characteristics, for example, the selected product category (Thirumalai & Sinha, 2005; Wan et al., 2016; Nguyen et al., 2019), distribution channel (Murfield et al., 2017), and the delivery carrier (Esper et al., 2003; Tokar et al., 2020). Findings from this literature confirm that customers expect a reliable, fast, and timely delivery service and express preferences based on the perception of the expected delivery service quality.

More recently, the literature has studied how last mile delivery operational performance influences final customers' experience. Traditionally the literature focused on delivery lateness (Rabinovich & Bailey, 2004). However, recent research has expanded to delivery earliness as another operational factor that positively shapes customers' experience. Specifically, Salari, Liu, and Shen (2022) found that customers are sensitive to the retailer's promise of real-time delivery time. Akturk et al. (2022) emphasized the importance that not only a late delivery negatively influences customer outcomes, but also an early delivery positively impacts customers outcomes. These conclusions are supported by a body of literature suggesting that early delivery is a gain for customers and late delivery a loss for customers (Tereyagoglu et al., 2018). In this study, we consider delivery punctuality as an additional factor influencing customers' outcomes. In the context of this study, delivery punctuality occurs when the delivery is within the promised time delivery window (Boyer et al., 2009; Mangiaracina et al., 2019). Studying punctuality in addition to earliness and lateness is important from an operational perspective and from a customer perspective. Crowdshipping platforms plan deliveries based on the delivery time window. An early

delivery, as late delivery, signifies a deviation from the order fulfillment process, and inaccuracy in operations that in the B2B context often results in inefficient operations and dissatisfied customers (Peng & Lu, 2017). From a customer perspective, theory and prior studies suggest that early deliveries do not have a strong positive impact on customer satisfaction, not as strong as the negative impact of late delivery. Indeed, (Mittal et al., 1998) describe the asymmetric effect of positive vs. negative attribute-level performance on customer outcomes, finding that customers remember less a positive performance as compared to a negative one. In addition, in last mile delivery context, (Mao et al., 2019) found that a late delivery has a much stronger negative effect on customers' outcomes than the positive effect of an early delivery. Thus, we examine how the delivery time performance within the delivery window influences the customer experience and outcomes. The first research question is RQ1: *How does delivery time performance impact customers' outcomes?*

While meeting the delivery window is crucial for customers' outcomes (Boyer et al., 2009), setting customers' expectations before the delivery takes place is equally important (Cui, Sun, et al., 2020). Customers may anchor their expectations relative to promised time (Salari et al., 2022). In this sense, the delivery window's length–the service window's length (Stroh et al., 2022), influences customers' expectations. Indeed, a short delivery window is more convenient for consumers than a long one, as consumers must spend less time at home waiting for the delivery (Nguyen et al., 2019). This is gorunded on Berry, Seiders, and Grewal (2002)'s logic on service convenience relative to consumers' time and effort. However, a longer delivery window creates flexibility in route planning and minimizes logistics costs for firms (Campbell & Savelsbergh, 2005; Boyer et al., 2009). In a similar fashion, the type of service the customer requests also impacts the expectations. For example, the request of an expedited delivery service, or the request

to choose a specific delivery window (sort of a delivery appointment), will shape customers' expectations. And the retailer's shipping speed promise contributes to shaping the expectations (Cui, Sun, et al., 2020). Thus, it is important to understand how the delivery window and the type of requested service shape customers' expectations relative to the delivery and the effect of the delivery performance relative to the expectation. Thus, the second research question of this study is RQ2: *How do customers' expectations moderate the relationship between delivery time performance and customer outcomes?*

We inform these research questions by adopting the key tenets of expectancy disconfirmation theory (EDT) and prior literature on service convenience and service experience. EDT states that customer outcomes result from the difference between the expectations of the provided service and the evaluation of the actual performance (Oliver, 1977, 1980). Based on this theoretical approach, we hypothesize that a decrease in operational performance will negatively affect customer outcomes and that customers' expectations moderate the relationship. While EDT typically focuses on the interplay between ex-ante expectations and ex-post evaluation, this study also extends EDT by assessing the role of time perception and service affect expectations and performance.

Similarly to prior literature that investigated customers' outcomes over a period of time (De Vries et al., 2016; Mao et al., 2019), we empirically investigate the research questions using a dataset comprising customers' transaction data over a three-month period. Research often relies on a multi-period dataset reporting transactional data to study customer behaviors over time. This is important because customers form expectations on a service delivery through accumulated experience (Zohar et al., 2002), because customers select retailers based on their past shopping experiences (Cui, Li, et al., 2020), and because customers' sensitivity to changes in delivery

service quality is contingent upon their experience with the service (M. Fisher et al., 2019). Specifically, we retrieved an archival dataset from a Fortune 100 retailer comprising approximately 14 million transactions with 5 million customers over three months.

The study aims to offer theoretical and managerial contributions. First, this study extends the literature on last mile delivery and addresses recent calls for empirical research on assessing actual customers' behaviors (Ketchen et al., 2021). Second, this study extends EDT by identifying the time boundary conditions affecting customers' outcomes. Finally, investigating such boundary conditions holds important consequences at the operational level. First, although prior literature investigated the key role of labor planning on crowdshipping last mile delivery systems to deliver on-time within guaranteed delivery time windows (Fatehi & Wagner, 2022), understanding customers' reaction to the delivery performance relative to the length of the delivery window may provide additional insights to the operations planning. Indeed, customers form expectations relative to the provided service quality (Thirumalai & Sinha, 2005). Hence, a late delivery that occurred during a one-hour delivery window may not have the same impact as a late delivery during an eight-hour delivery window. Thus, based on the effect of customers' expectations relative to the delivery service, retailers may relax the planning of the delivery system to adapt to customers' expectations.

Literature review: Consumers' challenges in last mile delivery

The second research stream refers to the literature on logistics service quality and physical distribution service quality. This research has been typically related to last mile delivery because it identifies the crucial elements of product fulfillment service quality that impact customer outcomes (Nguyen et al., 2019). This body of literature has used primarily empirical approaches to extensively investigate the logistics service quality dimensions of the order fulfillment process,

namely the operational dimension, the economic dimension, and the relational dimension (J. Mentzer et al., 1999; Stank et al., 2003; Collier & Bienstock, 2006b).

The operational dimension of logistics service quality is crucial for a successful delivery, and include delivery aspects such as number of shipping options, on-time and fast deliveries, as well as warehousing aspects such as product availability, ability to track orders, and product handling (Rabinovich, 2004; Xing & Grant, 2006; Ambra et al., 2021). Within this dimension, research has first identified the key factors that constitute the operational dimension of physical distribution service quality, then investigated the impact on retailer and customer outcomes. Specifically, Rabinovich and Bailey (2004) found that inventory availability, timeliness, and service reliability are driven by several determinants sorted depending on the e-retailer and transaction-specific attributes, whereas Stewart and Chase (1999) found that the human error is a core element in determining the quality of service quality. Consequently, physical distribution service quality affects operational and financial performance of the e-retailer (Rabinovich & Evers, 2003; Rabinovich, 2004; Rabinovich et al., 2007, 2008), as well as customer's outcomes, such as satisfaction, referral, repurchase intentions, and loyalty (Esper et al., 2003; Davis-Sramek et al., 2008; Rao, Goldsby, et al., 2011; Rao, Griffis, et al., 2011; Thirumalai & Sinha, 2011; Griffis, Rao, Goldsby, Voorhees, et al., 2012), customer's intention to return the item (Rao et al., 2014). The literature has also addressed the key role of product type in segmenting product fulfillment based on customers' expectations toward convenience, shopping, and specialty goods (Thirumalai & Sinha, 2005, 2009; Nguyen et al., 2018).

The economic dimension in last mile delivery refers to the price paid for the fulfillment service (e.g., shipping and handling charges) (Rabinovich & Bailey, 2004). Recently, this dimension has received more attention from the literature due to shifting customer's expectations

relative to paying for shipping fees (Gümüş et al., 2013). Ma (2017) found shipping charges to be moderating the perceived delivery time service quality, with greater satisfaction and purchase intention when customers achieve an ideal combination of long service time and free shipping or short delivery time paying charges. Nguyen et al. (2019) investigated customers' preferences toward delivery options, finding that delivery fees are the most important aspect impacting customers' purchasing behaviors. Similarly, Lewis, Singh, and Fay (2006), Barker and Brau (2020), and Tokar et al. (2020) studied the partitioning of delivery fees from the product's price. They found that partitioning affects customers' prepurchase intentions and basket size. Indeed, customers are not willing to pay shipping fees even when a carrier with greater performance is provided.

Finally, the relational dimension includes aspects of the interaction occurring in the order fulfillment process between the focal firm and the customer (Davis-Sramek et al., 2008). While this dimension is deemed important because customers are replacing the point-of-purchase salesperson with service across the last mile (Daugherty et al., 2019; Peinkofer et al., 2020), the relational dimension in last mile delivery has received only limited attention. Ta, Esper and Rossiter (2018) investigated the impact of crowdsourced driver's ethnicity disclosure on customers' outcomes, discussing social topics such as driver discrimination and the driver's sacrifice in last mile delivery (Esper, 2021). Instead, Castillo et al. (2022) looked at the relationship in the last mile using the drivers' perspective, specifically focusing on how customers' tipping influences drivers' performance.

Overall, this literature identifies the key importance of providing customers with high levels of logispriortics service quality but overlooks the increasing customers' expectations toward last mile delivery. Indeed, customers have expressed new preferences for deliveries, which harm

the efficiency and effectiveness of the service (Hübner et al., 2016; Daugherty et al., 2019). Consumers expect on-time deliveries with a reduced fulfillment time at their convenience, pushing the time performance of the delivery service to achieve both speed and punctuality (Gawor & Hoberg, 2019; Mangiaracina et al., 2019). Delivery speed, or the interval between the customer order and the delivery, is an essential service component of customer satisfaction (M. Fisher et al., 2015). Retailers offer several delivery speed options to win online customer orders, such as expedited delivery, same-day delivery, and next day delivery (Ishfaq et al., 2016; Peinkofer et al., 2020), increasing the pressure on an already-thin profit margin fulfillment service that is yet indispensable, especially for large retailers (Dayarian & Savelsbergh, 2020; Kammerer et al., 2020; Stroh et al., 2022). Delivery punctuality, defined as receiving the package within an established delivery time lapse (Mangiaracina et al., 2019), is highly valued by consumers (Xing et al., 2010; Nguyen et al., 2019), but it becomes problematic due to customized delivery windows, which is a constraint when planning last mile deliveries (Lim & Winkenbach, 2019; Praet & Martens, 2020). Hence, the second research gap relates to addressing customers' increased expectations of punctual and fast deliveries.

Expectancy Disconfirmation Theory

According to Expectancy Disconfirmation Theory (EDT), individuals assess the quality of a service performance by comparing prior expectations with the actual performance (Oliver, 1977, 1980). An individual's expectations correspond to the subjective probability or beliefs relative to the attributes of service before the experience (Olson & Dover, 1979). Upon experiencing the service performance, individuals assess whether the experience met or unmet the expectations. A disconfirmation occurs when the level of service does not match the level of expectations (Oliver, 1977). A disconfirmation is positive (negative) when the perceived performance (does not)

exceeds the expectations. In the former, individuals positively assess the service performance, and satisfaction follows, whereas the second results in individuals being dissatisfied (Anderson & Sullivan, 1993; Smith & Bolton, 2002; Hess et al., 2003). EDT has been widely adopted in operations management literature investigating how customer outcomes are influenced by expectations relative to the service quality (Zohar et al., 2002; Ho & Zheng, 2004; Venkatesh et al., 2010; Li et al., 2013; Chen et al., 2018; Dixon & Thompson, 2019; Kokkodis et al., 2022), the order fulfillment process (Rao, Griffis, et al., 2011; Rao et al., 2014; Peinkofer et al., 2016; Serkan Akturk et al., 2018), and more specifically to the last mile delivery service (Akturk et al., 2022; Salari et al., 2022). Overall, this literature finds that not only expectations disconfirmation shape customer satisfaction, but also customers' behaviors, specifically purchasing behaviors, including repurchase intentions, decrease in order frequency and order size (Rao, Griffis, et al., 2011).

EDT is grounded on the premise that individuals follow a chronological process when assessing service performance relative to prior expectations (Oliver, 1980). This process comprises of four sequential steps leading to customers' repurchase behaviors (Bhattacherjee, 2001). First, the customer forms an expectation relative to the delivery service. This process is characterized by information asymmetry on the level of delivery service (Rabinovich & Bailey, 2004; Kokkodis et al., 2022). Customers expect the delivery service to follow standards of reliability and dependability, such as timeliness and speed (Mentzer et al., 1999; Li et al., 2005; Akturk et al., 2022). Thus, upon ordering from a retailer, customers form the expectation to quickly receive the order on-time. This expectation is based on information asymmetry because customers estimate the level of delivery service only on the information available, such as the promised delivery time and real-time tracking data (Rao et al., 2014; Akturk et al., 2022; Salari et al., 2022), or prior experience with the service provider (Zohar et al., 2002).

Second, the customer accepts and experiences the delivery service, and forms perceptions about the performance and the level of service quality (Bhattacherjee, 2001). Thus, customers proceed with the home delivery request to the retailer, and experience the actual delivery performance. The delivery performance is influenced by many aspects related to the delivery driver (Bode et al., 2011; Ta et al., 2018; Awaysheh et al., 2021; Liu et al., 2021; Castillo et al., 2022), to the challenges of the context in which it is performed (Amling & Daugherty, 2018; Deng et al., 2021; Merkert et al., 2022; Rose et al., 2022), or to the characteristics of the delivery (Boyer et al., 2009; Nguyen et al., 2019; Barker & Brau, 2020; Akturk et al., 2022; Stroh et al., 2022). Despite the multitude of factors that may influence the performance, customers form perceptions only based on the experienced level of service quality, i.e., customers often consider late delivery service failures (Rao, Griffis, et al., 2011).

Third, the customer uses prior expectations as a frame of reference to evaluate the performance through a comparative judgment that results in confirmation or disconfirmation (Oliver, 1980). A failure to meet customers' expectations results in a disconfirmation, and the value of the service is lowered (Bolton, 1998). In last mile delivery, a disconfirmation typically occurs when an order is late, when the order fulfillment process is slow, and when the shipping cost does not match the delivery performance (Gümüş et al., 2013; Akturk et al., 2022).

Fourth, the customer forms a satisfaction based on the disconfirmation level, and repurchase intentions follows (Bhattacherjee, 2001). Specifically, satisfaction results from confirmation or positive disconfirmation, whereas dissatisfaction comes from negative disconfirmation. Satisfaction is the building block of repurchase intentions because a satisfied customer is likely to continue service use (Anderson & Sullivan, 1993; Oliver, 1993). Hence, a

delivery that disconfirms prior expectations will likely result in dissatisfaction and the customer will be less likely to repurchase from the retailer.

Hypotheses development

To formulate the hypotheses of this study, we integrate EDT with the logic of service convenience and service experience literature. The outcome of interest is customer outcomes, which refers to the measurable results that a customer can perceive in terms of value generated by the service provision (Hennig-Thurau et al., 2006; Rao, Goldsby, et al., 2011). In this study, we capture customer outcomes in terms of customer satisfaction and repurchase behaviors. These two customer outcomes are of key importance for an online retailer and service provider because they assess the value customers perceive during the online shopping experience and the delivery service (Rao, Griffis, et al., 2011). Prior literature has also adopted satisfaction and repurchase behaviors as outcomes of interest when studying customers' perceived service quality (Davis-Sramek et al., 2008; Buell et al., 2010). Figure 17 provides an overview of the theoretical model.

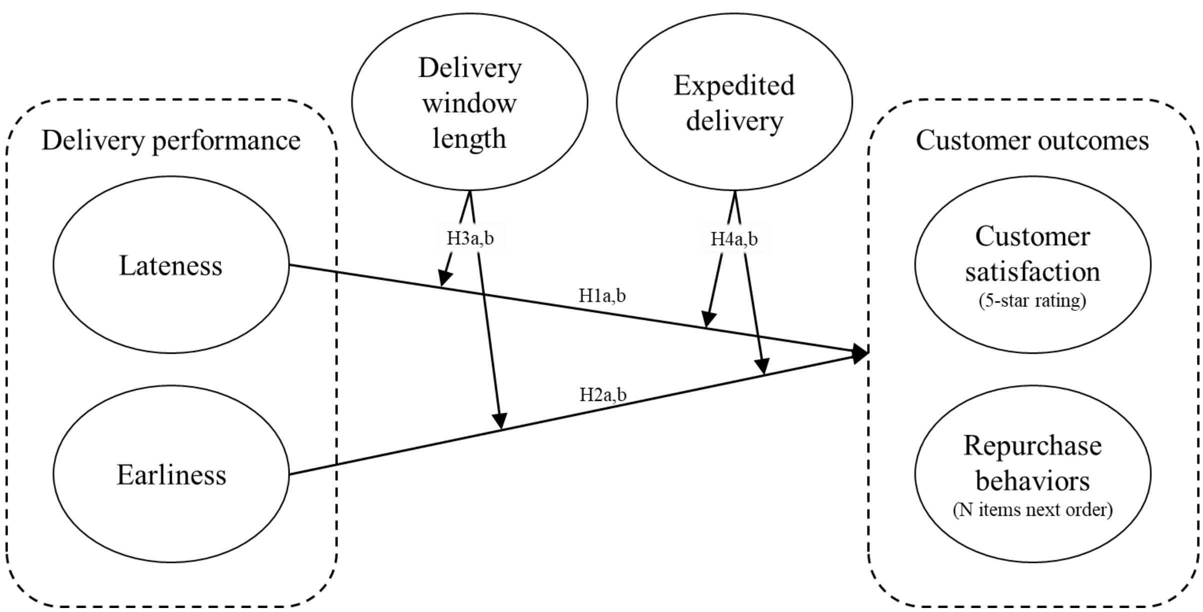


Figure 17 – Theoretical model essay 3

The impact of delivery time performance on customer satisfaction

Customer satisfaction is defined as the overall assessment of the events during the service experience (Anderson et al., 1994). In online retailing, customer satisfaction has often been associated with the customer rating of the experience. Customer rating is a popular online tool to measure customer satisfaction and reflect the value the customer perceived from the delivery experience (Gordon et al., 2015; Alexander et al., 2021; Akturk et al., 2022). Hence, similar to extant studies, this study assumes that the reported ratings of consumers reflect their satisfaction after consumption (Chen et al., 2018). Upon placing an online order and requesting the delivery service, customers form the expectation to receive the order on-time. Once the delivery driver consigns the order, the customer evaluates the service experience through a confirmation or disconfirmation of the expectations. A late (early) order results in a negative (positive) disconfirmation.

A negative disconfirmation occurs because customers perceive a service failure in the delivery service (Rao, Griffis, et al., 2011). Hence, a negative disconfirmation will result in customer dissatisfaction. This aligns with prior studies that a late delivery decreases customer satisfaction (Lee & Whang, 2001; Boyer & Hult, 2006; Fisher et al., 2019). Finally, a dissatisfied customer will rate less the overall experience, assigning lower online ratings to the delivery service. Prior literature identifies a difference between positive and negative disconfirmation because a customer's loss has a much stronger negative impact on outcomes than a customer's gain (Mittal et al., 1998; Tereyagoglu et al., 2018). Hence, an early delivery – a delivery occurring before the beginning of the delivery window, represents a gain for customers, who perceive the delivery system outperforming the expected delivery service (Brown et al., 2008). This is likely because 1) customers perceive quicker gratification from online purchases (Balasubramanian,

1998), and 2) customers perceive the delivery process as being less risky and more efficient than expected (Rabinovich & Bailey, 2004). Hence, in instances of early delivery, customers will experience a positive disconfirmation followed by satisfaction and will provide a higher rate to the experience. Hence, we expect:

Hypothesis H1a: *Lateness is negatively associated with customer satisfaction.*

Hypothesis H2a: *Earliness is positively associated with customer satisfaction.*

The impact of delivery time performance on customer outcomes with the retailer

Repurchase behaviors is defined as customer's attitude to engage in future repurchase behavior towards a retailer (Seiders et al., 2005). EDT predicts that satisfied customers experiencing a confirmation or a positive disconfirmation will engage in repurchase behaviors with the retailer (Bhattacherjee, 2001). This is because an experience exceeding or meeting the expectations will reduce the asymmetry of information relative to the delivery time performance. Indeed, in such occurrences, customers update their expectations regarding future transactions and service provisions through a series of anchoring and adjustment processes, which leads customers to adjust their expectations based on the successful service provisions (Hogarth & Einhorn, 1992; Rao, Griffis, et al., 2011). Thus, a customer experiencing a satisfactory service provider is expected to continue repurchasing from the retailer. In contrast, a negative disconfirmation results in reduced repurchase from the retailer. These arguments have been established in prior service quality literature (Bitner, 1990, 1995; Smith et al., 1999; Davis-Sramek et al., 2008; Wan et al., 2016).

In the context of this study, customer repurchase behaviors manifests through customer's future spending in terms of order size. Future spending has traditionally been linked to customer repurchase behaviors (Guo & Liu, 2023), used as the primary dependent variable in prior literature investigating delivery operations (Rao, Griffis, et al., 2011; Bhan & Anderson, 2023), and

theorized as a key customer outcome in sharing economy contexts (Luo et al., 2021). A change in order size reflects a change in the value of the relationship between the customer and the retailer (Lewis et al., 2006; Rao, Griffis, et al., 2011). The literature typically studies a change in order size relative to the shipping fees applied to online orders, investigating customers' behaviors relative to adding items to the total order to avoid shipping fees (Khan et al., 2009; Leng & Becerril-Arreola, 2010; Han et al., 2022). Following this logic, in this study, a change in order size corresponds to an increase or decrease in the number of items included in the order.

Delivery time performance is a key service quality aspect influencing customers' behaviors toward the retailer. A delivery service that negatively disconfirms customers' expectations will likely decrease the perceived value of the transaction (Lewis et al., 2006). Hence, the customer will likely reduce the intention to patronize the retailer because of the service quality uncertainty (De Vries et al., 2018). In contrast, a positive disconfirmation will hold the opposite result of increasing order frequency and order size because customers perceive less uncertainty, adjust their expectations to receive their shopping within the expected amount of time, and will not seek an alternative. Hence, we expect:

Hypothesis H2a: *Lateness is negatively associated with repurchase behaviors.*

Hypothesis H2b: *Earliness is positively associated with repurchase behaviors.*

The moderating role of customer's expectations

Customers form expectations of the delivery service based on available information and prior experience (Bhattacherjee, 2001). However, these expectations depend on the subjective perception of waiting time (Zohar et al., 2002). Hence, expectations differ based on the length of time customers wait for the service provision, being a short wait or a long one (Zohar et al., 2002). In last mile delivery, customers often face variance in waiting time. In this study, we do not refer

to waiting time as the time for the order to be fulfilled (Akturk et al., 2022), rather as the length of the delivery window. Studies have looked at customers' responses to the width of the delivery time slot (Amorim et al., 2020). Customers prefer greater precision (shorter delivery window) to increase the time convenience of the service provision (Berry et al., 2002; Boyer et al., 2009; Nguyen et al., 2019).

When the service provider sets the delivery window length, customers form expectations on the level of service quality, because the window indicates the guaranteed delivery time (Stroh et al., 2022). A longer delivery window is likely to increase the chances of receiving the order on-time (i.e., within the delivery window). Given the aim of the customer to receive the order on-time, a longer delivery window will create the expectation that the order will likely be delivered within the targeted window (Grout, 1998). A longer delivery window reduces uncertainty relative to the waiting time, increasing the predictability of the delivery and increasing the customer's perception of control (Bitran et al., 2008). Providing customers with information about the wait helps manage expectations (Bitran et al., 2008). This has recently been investigated in last mile delivery: Research found that including a feed of real-time delivery enhances customer outcomes (Salari et al., 2022). In some instances, major retailers also allow customers to pick a preferred time slot for delivery (Agatz et al., 2021). In contrast, a shorter delivery window may reduce the expectations that the order will be on-time, increasing uncertainty on the wait time, thus decreasing customers' perception of control over correctly estimating the waiting time (Donohue et al., 2020).

Consequently, different effects hold for the moderation of delivery window length on the relationship between delivery time performance and customer outcomes. A longer delivery window will increase expectations of an on-time delivery, by decreasing the uncertainty of late delivery. An increase in expectations results in a stronger disconfirmation in case of late deliveries

because of the larger difference between expectations and service performance. However, higher expectations reduce the difference between prior expectations and performance in the case of an early delivery. Hence, while delivery window length magnifies the impact of late deliveries on customer outcomes, it mitigates the positive effect of early deliveries on customer outcomes. Shorter delivery windows bear the opposite effect, mitigating the effect of a late delivery because of lower expectations while magnifying the effect of on-time or early deliveries on customer outcomes. Thus, we expect that:

Hypothesis 3a,b: *Delivery window length moderates the relationship between delivery time performance and customer outcomes, such that a longer delivery window will a) magnify the negative relationship between lateness and customer outcomes, but b) mitigate the positive relationship between earliness and customer outcomes.*

Another element impacting customer's expectations is the delivery service the customer request to the retailer. In last mile delivery, customers may require a fast delivery, which refers to the expedited delivery for time-definite and sensitive services (Peinkofer et al., 2020). Given that customers are sensitive to the delivery time (Fisher et al., 2019), faster delivery will provide customers with the benefit of reducing uncertainty relative to the waiting time. Hence, for expedited deliveries, customers may be tolerant relative to delays and disclose greater satisfaction and repurchase behaviors when the service meets or exceeds expectations. However, the request for this additional service impacts the delivery expectations because customers are usually requested to pay an extra fee. Prior literature found that customers are reluctant to pay delivery fees and that delivery fees are an important determinant of customers' purchasing behaviors (Ma, 2017; Nguyen et al., 2019; Barker & Brau, 2020). For example, the shipping fee is a key factor driving the conversion rate in online retailing, with many consumers leaving their shopping cart

upon requesting a shipping fee (Leng & Becerril-Arreola, 2010; Akturk & Ketzenberg, 2022; Han et al., 2022). In the context of this study, customers can require an expedited delivery service by paying additional fees. The effect is to increase customers' expectations: as customers pay an additional premium to receive the order faster, they also increase their expectation to receive the order in a timely fashion. Hence, we expect that:

Hypothesis 4a,b: *Fast shipping moderates the relationship between delivery time performance and customer outcomes, such that a faster shipping will a) magnify the negative relationship between lateness and customer outcomes, but b) mitigate the positive relationship between earliness and customer outcomes.*

Methodology preview

We empirically test these hypotheses through a multi method design approach, including econometric analysis of archival data and scenario-based experiments (Golicic & Davis, 2012). We chose this research design for many reasons. First, while empirical research using archival data ensure generalizability of the results, a common limitation is presence of endogeneity (Lu et al., 2018; Miller et al., 2020). In contrast, experiments present the advantage of a controlled scenario, in which the researcher manipulates an independent variable to observe its effect on a dependent variable (Lonati et al., 2018). Second, the nature of the dependent variable *online rating* may raise questions relative to the actual representativity of the customer experience. For example, prior literature discussed that only a proportion of the customer population leaves feedback, hence including a reporting bias (Karamana, 2021). Finally, the dependent variables capturing customer outcomes with the retailer reflect the actual purchasing behavior of the customer. While this is preferable, prior literature has demonstrated the importance of capturing customers' intent toward their repurchasing behaviors (Davis-Sramek et al., 2008; Wan et al., 2016). Hence, Study 1 refers

to the econometric analysis of a dataset retrieved from an online retailer (hereafter named *Alpha*). Study 2 refers to the scenario-based experiment.

Study 1 – Data and econometric models

Data description

We empirically investigated these hypotheses by compiling a dataset from multiple sources. First, we retrieved delivery operational data and driver outcomes data from a Fortune 100 retailer (hereafter called *Alpha*). The retailer has launched its white-label crowdshipping platform, which performs home deliveries from the retailer's stores using crowdsourced drivers. Upon being pinged from the retailer for home deliveries, the crowdshipping platform broadcasts the offer for the delivery task through the Driver App, where the offer is visible to a set of drivers assigned to a driver zone[11]. The offer includes details related to the compensation, the pre-tipping, the delivery type, the number of orders included in the delivery task, the total of miles that a driver would drive to perform the task (i.e., from the store to the final drop-off), as well as the store and customers' address. Upon accepting the task, the Driver App provides GPS navigation instructions to arrive at the pick-up point and drop-off the orders at the customer's destination.

Alpha shared a raw dataset covering three months (February to April 2022), and containing approximately ~14.3 million customers' orders delivered by ~173k drivers. The dataset includes detailed information about the characteristics of the order and the customers. The dataset reports several time stamps capturing the hour, minute, and second of the delivery operations for the order i, including order delivered ($orde_j$), which refers to the timestamp of when the driver delivered the order, and delivery time window end (de_j), which refers to the planned delivery time for a

[11] As common in crowdshipping (Guda & Subramaniana, 2019; Tripathy et al., 2022), the platform assigns drivers to a driver zone, which serves the stores located in at least one zip code (i.e., the same zone may serve multiple zip codes).

specific order. This dataset also includes the delivery type for each order (expedited or standard), the geo coordinates (latitude and longitude) of each drop-off location, as well as information relative to the store that dispatched the delivery task, including store id and complete address.

Second, we manually retrieved publicly available data from *Alpha*'s website about stores that dispatched a delivery task. Specifically, we downloaded data on stores' id, type, and exact address. *Alpha* sorts stores dispatching a delivery task into different[12] types, based on the dimension of the store. The exact address includes street name and number, 5-digit zip code, town, and state. Next, we built a custom program to extract the geographical coordinates (i.e., longitude and latitude) from each store address. Following current literature (Belo et al., 2014; Belenzon et al., 2020; Barrios et al., 2022), we employed Google Maps Geocoding application programming interface (API), which is a Google Cloud Platform powered by Google Enterprise API, that converts between addresses and geographic coordinates. The platform library provides an XML file to automate the search for coordinates from the specified address and guarantees complete integration with Microsoft Excel. Among the advantages of this API is the correction of small variations in spelling (misspelling) between the manually retrieved and the actual address, allowing to find the geocoordinates when there is a close match (Belenzon et al., 2020). We manually checked a sample of these conversions to ensure the precision of the conversion, without finding any unreasonable matching (Belenzon et al., 2020). Finally, we matched this and *Alpha* datasets through store id. Hence, using the geocoordinates of drop-off locations (from *Alpha*) and of store locations (from Google), we computed the store-drop-off distance, expedited in miles, using *geodist* function in STATA17. This function computes ellipsoidal distances (i.e., "the length of the shortest curve between two points along the surface of the mathematical model of the earth

[12] For confidentiality reasons, the number of drivers, stores, zip codes, delivery zones, and store types are not disclosed.

WGS 1984 datum" (Picard, 2022) – the same used by Google Earth) using Vincenty (1975)'s equation (Picard, 2022) A limitation is that geodist computes the actual distance, not the travel distance. However, the nature of the dataset being limited to local store-to-home deliveries allows to reasonably assume that g_n and travel distance are similar and strongly correlated.

Data cleaning

Before performing the main analyses, we cleaned the dataset following best practices of recent literature investigating similar contexts (Farber, 2015; Miao et al., 2022). First, we removed incomplete or erroneous observations and outliers for all the variables of interest. Specifically, we removed (1) orders with missing timestamps, (2) erroneous observations with total travel distance longer than 21 miles[13] (3) erroneous observations of ordered delivered before 6 AM and after 10 PM[14]. In the end, 13,927,531 observations were retained as final sample. Among these observations, we could observe only a subsample of orders which received a star rating from customers (n = 1,363,440). Theory and literature suggest that this is likely because customers initiate an information flow (i.e., post-service survey) only when extremely dissatisfied or extremely satisfied (Anderson, 1998; Chen et al., 2011; Taken Smith, 2012), especially when the experience is negative (Mittal et al., 1998). We address potential sources of endogeneity for sample bias prior to the main analysis. The unit of analysis of this study is at the single-order level.

Variables construction

We computed the outcome variables and predictors following theory, prior literature, and best practices. Table 16 reports the descriptive statistics, and Table 17 reports the correlations.

[13] This threshold result from symmetrically winsorized distance traveled at 99%, which produces an average distance traveled of 4.3 miles (SD = 4). This aligns with prior literature investigating similar context (i.e., crowdshipping), for example, (Miao et al., 2022) average trip distance was 11km (~7 miles), and (Castillo et al., 2022)'s netnography report drivers performing between 7 and 10 miles.
[14] *Alpha* provides the delivery service from 7 AM to 9 PM.

Table 16 – Descriptive Statistics

	Variable	Mean	SD	Min	Max	Source
(1)	Customer rating^{+}	4.66	1.00	1	5	(Alexander et al., 2021)
(2)	Repurchase behaviors*	14.49	19.6	0	90	(Luo et al., 2021)
(3)	Lateness	7.33	22.1	0	124.89	(Thürer et al., 2020)
(4)	Earliness	1.1	3.48	0	20.38	(Akturk et al., 2022)
(5)	Window length	2.84	4.26	0.51	22.25	(Boyer et al., 2009)
(6)	Expedited	0.06	0.23	0	1	(Peinkofer et al., 2020)
(7)	Order size	19.67	22.34	1	1212	(Liu et al., 2021)
(8)	Distance from store*	3.48	2.25	0.29	9.64	(Akturk & Ketzenberg, 2022)
(9)	N late previous orders	0.63	1.9	0	68	(Luo et al., 2021)
(10)	N early previous orders	0.36	2.49	0	84	(Luo et al., 2021)
(11)	Subscription	0.69	0.46	0	1	(Wagner et al., 2021)
(12)	Unattended	0.92	0.27	0	1	(Hübner et al., 2016)

Note: $^{+}$ ln transformed prior to the final analysis * The variable was symmetrically winsorized at 99%.

Table 17 – Correlations

	Variable	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
(1)	Customer rating	1.00											
(2)	Repurchase intentions	0.01***	1.00										
(3)	Lateness	-0.11***	-0.09***	1.00									
(4)	Earliness	0.02***	0.05***	-0.10***	1.00								
(5)	Window length	-0.01***	-0.17***	0.33***	-0.14***	1.00							
(6)	Expedited	-0.01***	0.04***	-0.07***	-0.01***	-0.09***	1.00						
(7)	Order size	0.02***	0.33***	-0.16***	0.06***	-0.32***	0.03***	1.00					
(8)	Distance from store	0.01***	-0.14***	0.26***	-0.24***	0.32***	-0.12***	-0.26***	1.00				
(9)	N late previous orders	0.01***	0.05***	-0.07***	0.25***	-0.09***	0.03***	0.02***	-0.19***	1.00			
(10)	N early previous orders	-0.01***	-0.04***	0.06***	-0.03***	0.07***	-0.02***	-0.06***	0.06***	0.04***	1.00		
(11)	Subscription	0.03***	0.28***	-0.16***	0.12***	-0.28***	0.05***	0.23***	-0.24***	0.19***	0.04***	1.00	
(12)	Unattended	0.01***	-0.04***	0.04***	-0.01***	0.09***	-0.03***	-0.10***	0.08***	0.00***	0.02***	-0.05***	1.00

Note: * $p < .1$ ** $p < .05$ *** $p < .01$

Customer satisfaction was operationalized as the customer feedback in the form of 5-star rating after the delivery was completed. Upon delivery completion, *Alpha* emails the customer the confirmation of the completed delivery and asks a 5-star rating feedback with the prompt: "How was your experience?" To support the validity of this operationalization, we recall the abundance of operations management literature that have used a similar proxy for customer satisfaction (Rao, Goldsby, et al., 2011; Alexander et al., 2021; Akturk et al., 2022; Ta et al., 2023). In our dataset, similar to prior literature (Akturk et al., 2022), customer satisfaction (hereafter rating) presents a

J-shaped distribution (Figure 18), with 73,002 orders with 1-star rating, 18,378 with 2-star rating, 28,656 with 3-star rating, 61,558 with 4-star rating, and 1,181,846 with 5-star rating. As expected, the two most frequent rating is 1-star and 5-star ratings, confirming the polarity of rating (i.e., customers rate the experience when extremely dissatisfied or extremely satisfied) (Anderson, 1998).

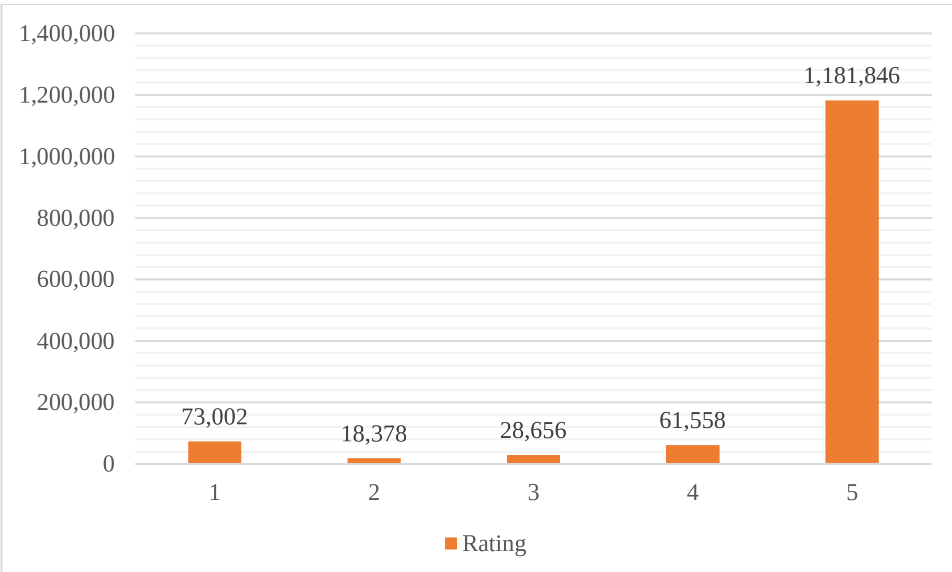


Figure 18 – Distribution of Customer Rating

Repurchase behaviors was operationalized as the total number of items the customer purchased on the next order (Rao, Griffis, et al., 2011; Luo et al., 2021). This operationalization capture the actual customers behavior on their next purchase. A limitation of this variable refers to the limited time span of the dataset, which does not allow observing the next order's actual time for all the orders, specifically those placed close to the end of the dataset. In these instances, common practice suggests treating the end of the dataset as the last customer order (De Vries et al., 2018). This operationalization is required to account for the censored nature of the dataset (Hosmer & Lemeshow, 1999). A second limitation is that this dataset only capture deliveries to the customer, hence does not observe whether the customer repurchase from the same retailer but

on a different channel. Despite these limitations we deemed this operationalization appropriate as it groups on theory and prior research (Rao, Griffis, et al., 2011; Bhan & Anderson, 2023). A greater value indicates a higher repurchase behaviors as customers increase the total number of items purchased and delivered from the retailer.

The focal predictors were operationalized following prior literature. Specifically, *Lateness* is computed as the time elapsed between the actual delivery time stamp and the time stamp indicating the end of the delivery window (Thürer et al., 2015, 2020). *Lateness* for the order i was computed as follows:

$$Lateness_i = \max(0, D_i - de_i)$$

Where D_i indicates the delivery time stamp of order i, and de_i is the delivery window end time stamp of order i. Positive values indicate a late delivery, whereas negative values indicate an on-time delivery. Similarly, *Earliness* is computed as the time elapsed between the time stamp indicating the beginning of the delivery window and the actual delivery time stamp (Akturk et al., 2022; Dayarian & Pazour, 2022). *Earliness* for the order i was computed as follows:

$$Earliness_i = \max(0, db_i - D_i)$$

Where db_i denotes the beginning of the delivery time window. Positive values indicate an early delivery.

The first moderator refers to delivery window length and is operationalized as the time elapsed between the beginning of the delivery window time stamp and the end of the delivery window time stamp. Delivery window length for the order i is computed as follows:

$$Delivery\ window\ length_i = db_i - de_i$$

Greater values indicate a longer delivery window. Finally, we operationalized expedited delivery as a dummy variable taking values of 1 if the customer requested an expedited delivery, 0 otherwise (Peinkofer et al., 2020).

We also included a set of control variables. First, we included order size, computed as the total number of items in order i. Order size has been found to service time and it is likely to affect customer expectations relative to delivery service based on how much they purchased (Rabinovich & Bailey, 2004; Liu et al., 2021). Second, we include *Distance*, computed as the distance, in miles, between the customer's delivery drop-off and the store that fulfilled the order. Prior literature found heterogenous customers' preferences relative to the omnichannel distribution (Lim et al., 2018; Agatz et al., 2021). The theory on service convenience suggests that less effort and lower time spent increase service convenience and so customer outcomes (Berry et al., 2002). For example, dissatisfied customers may decide to replace the delivery with the buy-online-pickup-in-store service (Dayarian & Pazour, 2022). Third, we included two variables capturing the cumulative number of late orders and early orders a customer experienced before the delivery took place. We included these control variables for two reasons. Theory suggests that customer form expectations also based on their prior experience (Bhattacherjee, 2001). In addition, recent theoretical developments in the context of sharing economy suggest that prior expectations influence customer outcomes (Luo et al., 2021). Third, we include *subscription*, a dummy variable indicating whether, at the time of order i, the customer pays a subscription for the provision of the delivery service. Many retailers offer subscription plans for the delivery service (e.g., Amazon Prime) (Caro et al., 2020). This is likely to influence not only customers expectations, because of the subscription fee, but also repurchase behaviors, because a customer that has paid for a subscription plan is more likely to engage in repurchase behaviors with the same delivery service

provider (Wagner et al., 2021). Finally, *Unattended* is a dummy variable indicating whether the delivery was unattended, that is the customer not being at home (Hübner et al., 2016). For unattended deliveries, the customer may develop lower expectations relative to the waiting time because the customer is not at home. Finally, we include a set of control variables related to the time of the day, *Afternoon*, a dummy variable taking one if the delivery was completed in the afternoon, zero otherwise, *Weekday*, a dummy variable taking one if the delivery was completed during the working week (Monday to Friday) (Choudhary et al., 2021), and fixed effects for the month.

Preliminary analysis

Following prior literature (Corbett et al., 2005; Hendricks & Singhal, 2014), we symmetrically winsorized at 99% repurchase behaviors to limit the effects of outliers on the final analysis. In addition, we computed the natural logarithm for customer rating to correct for distribution skewness (Andritsos & Tang, 2014; Akturk et al., 2022).

Endogeneity concerns

The purpose of this study is to investigate the impact of delivery performance on customer outcomes. Despite the effort to collect several control variables, upon consulting operations management literature, we identified a potential source of endogeneity due to omitted variable bias, which manifests when unobservable or unavailable factors affect both predictors and outcome variables (Ho et al., 2017; Lu et al., 2018; Mithas et al., 2022). Prior literature investigating similar research questions suggests that unobserved predictors related to fulfillment operations may affect the delivery performance (Rabinovich & Bailey, 2004; Akturk et al., 2022), and consistently impact customer outcomes, given that customers are inherently more sensitive to lead time in online channels (Lim et al., 2020). To overcome this challenge, we adopted a IV/2SLS

econometric approach, which ensures to identify the causal estimation for the effect of the focal predictors on the outcome variables (Ho et al., 2017).

In the first step of the IV/2SLS, we identified instrumental variables that meets the relevance and exclusion condition, per Wooldridge (2010). Indeed, an instrument must explain the suspected endogenous predictor (relevance condition), and affect the outcome variables only through the focal predictor being unrelated to the unobservable variables included in the error term (Wooldridge, 2016). The challenge presented in this context is to identify endogenous predictors that would uniquely affect each predictor (e.g., lateness but not earliness). For example, delivery density would appropriately impact both lateness and earliness of an order. Hence, following prior literature (Akturk et al., 2022), we instrumented lateness (earliness) with the subgroup average lateness (earliness) at the delivery zone level. It is plausible that delivery operations within a delivery zone are systematically affected by unobserved factors such as the level of traffic, road conditions, weather conditions, as well as stores' fulfillment operations. Following Akturk et al. (2022) arguments, the average lateness (earliness) could explain the variation of lateness (earliness), without impacting customer outcomes, given that it is unlikely that customers know the average lateness (earliness) for all deliveries completed within the delivery zone.

Estimation Models

Prior to estimating the first stage of IV/2SLS, we reviewed the distributions of the two endogenous predictors (lateness and earliness), which indicate that these two predictors follow a Poisson distribution. Thus, we followed Kamalahmadi et al. (2021) approach, which utilizes different estimation approaches between the first and second stage of IV/2SLS procedure. Specifically, in the first stage, we adopt a Poisson regression, whereas in the second stage we adopt a linear

regression for customer rating, and a Poisson regression for repurchase behaviors. We detail the regression models as follows:

2SLS-Stage 1. Run a Poisson regression for Lateness and Earliness on the instruments (average lateness and average earliness at delivery zone z), and control variables of vector X_i, as follows:

(19) $\Pr(Lateness_i = s|X_i) = Poisson(\alpha + \beta_1 \mu Lateness_z + \beta_2 \mu Earliness_z + \mathrm{BX}_i + \varepsilon_i)$

(20) $\Pr(Earliness_i = s|X_i) = Poisson(\alpha + \beta_1 \mu Lateness_z + \beta_2 \mu Earliness_z + \mathrm{BX}_i + \varepsilon_i)$

Where X_i is the vector of control variables, B is the vector of coefficients for the control variables, and ε_{id} are robust standard errors, clustered for driver id to mitigate the potential source of heteroscedasticity. BX_i was specified as follows:

(21) $\mathrm{X}_i = \beta_0 + \beta_1 order\ size_i + \beta_2 dist_d + \beta_3 cum\ late\ orders_i + \beta_4 cum\ early\ orders_d + \beta_5 subscription_z + \beta_6 unattended_i + \beta_{7is} store_{is} + \beta_{8im} month_{im} + \beta_{9iw} dow_{iw} + \beta_{10it} time_{it}$

The predicted values for lateness and earliness are $\widehat{Lateness}_i$ and $\widehat{Earliness}_i$ respectively.

2SLS-Stage 2. Run an OLS regression on customer rating of order i replacing $Lateness_i$ and $Earliness_i$ with $\widehat{Lateness}_i$ and $\widehat{Earliness}_i$, as follows:

(22) $\ln(\text{customer rating})_i = \alpha_0 + \alpha_1 \widehat{Lateness}_i + \alpha_2 \widehat{Earliness}_i + \mathrm{BX}_i + \varepsilon_i$

A Poisson regression for Repurchase behaviors of customer d for order $i + 1$, as follows:

(23) $\Pr(Repurchase_{di+1} = s|X_i) = Poisson\left(\alpha + \alpha_1 \widehat{Lateness}_i + \alpha_2 \widehat{Earliness}_i + \mathrm{BX}_i + \varepsilon_i\right)$

Study 1 – Results

We first presented statistical evidence relative to our claims for endogeneity. Then, we reported the testing of each hypothesis and a plot for each significant interaction effect.

Tests for endogeneity

Results from the first stage (Table 18 Models 1-2) report that average lateness (earliness) positively impact lateness (earliness), yet negatively effect the alternative delivery performance.

Table 18 – IV/2SLS results of the first stage

	(1) Lateness		(2) Earliness	
Order size	0.008^{****}	(0.000)	-0.009^{****}	(0.000)
Distance from store	0.088^{****}	(0.003)	-0.330^{****}	(0.002)
N late previous orders	-0.136^{****}	(0.007)	0.072^{****}	(0.001)
N early previous orders	0.224^{****}	(0.017)	-0.117^{****}	(0.004)
Subscription	0.019	(0.015)	-0.020^{****}	(0.006)
Unattended	-0.085^{****}	(0.015)	0.112^{****}	(0.006)
Month fe	YES		YES	
Day of the week fe	YES		YES	
Clock hour fe	YES		YES	
Store fe	YES		YES	
IV Lateness	0.102^{****}	(0.002)	-0.017^{****}	(0.001)
IV Earliness	-0.068^{****}	(0.008)	0.475^{****}	(0.002)
_cons	2.133^{****}	(0.165)	1.567^{****}	(0.066)
χ^2	17681.370		210013.378	
McFadden's pseudo R^2	0.076		0.218	
N	1394399.000		1394399.000	

Note: robust standard errors reported in parenthesis. $^{+}$ $p < 0.1$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.005$, **** $p < 0.001$

Results of endogeneity tests (Table 19) suggest that the instrument met the relevance and exclusion conditions. Average lateness strongly correlates with lateness (r = 0.121), but weakly correlates with customer rating (r = -0.022) and repurchase behaviors (r = -0.065). Similarly, average earliness strongly correlates with earliness (r = 0.247), but weakly correlates with customer rating (r = -0.005) and repurchase behaviors (r = -0.039). In addition, the instrumental variable for lateness also presents high F-statistics in the first stage (F = 44,375, p < 0.01), well over 10 (Staiger & Stock, 1997), to conclude that is a strong instrument (Wang et al., 2022). The same conclusion is drawn for the instrumental variable for earliness (F = 40,090, p < 0.01). As expected, we found evidence of endogeneity. Durbin-Wu-Hausman tests for customer rating (F = 753.17, p < 0.01) and repurchase behaviors (F = 10,178, p < 0.01) confirm the presence of endogeneity and the choice of an IV/2SLS approach (Dhanorkar & Siemsen, 2021).

Table 19 – Test for endogeneity

Instrumental variable	**Cragg-Donald Wald F-test**	**Lateness**	**Earliness**	**Customer rating**	**Repurchase behaviors**
Average lateness at delivery zone level	F = 44,375 p < 0.01	r = 0.121	r = -0.096	r = -0.022	r = -0.065
Average earliness at delivery zone level	F = 40,090 p < 0.01	r = -0.047	r = 0.247	r = -0.005	r = 0.039
		Durbin-Wu-Hausmann χ^2		753.17 p < 0.01	10,178.4 p < 0.01

Direct effects on customer outcomes

Results are reported in Table 20. H1a predicted a negative effect of lateness on customer satisfaction, whereas H2a a positive effect of earliness on customer satisfaction. Model 1 reports a negative and significant coefficient of lateness (β = -0.041, p < 0.01), supporting H1a, but a negative and significant coefficient of earliness (β = -0.023, p < 0.01), not supporting H2a. H1b predicted a negative effect of lateness on repurchase behaviors, whereas H2b a positive effect of earliness on repurchase behaviors. Model 2 reports a negative and significant coefficient of lateness (β = -0.086, p < 0.01), supporting H1b, but a negative and significant coefficient of earliness (β = -0.019, p < 0.01), not supporting H2b.

Table 20 – IV/2SLS second stage. Direct effects of lateness and earliness on customer outcomes

	(1) Ln(rating)		(2) Repurchase	
order size	0.000****	(0.000)	0.009****	(0.000)
distance from store	-0.000	(0.000)	0.019****	(0.001)
n late previous orders	-0.002****	(0.000)	0.010****	(0.001)
n early previous orders	-0.001	(0.001)	0.005*	(0.002)
subscription	0.024****	(0.001)	0.255****	(0.004)
unattended	0.000	(0.001)	0.034****	(0.003)
Month fe	YES		YES	
Day of the week fe	YES		YES	
Clock hour fe	YES		YES	
Store fe	YES		YES	
Lateness	-0.041****	(0.001)	-0.086****	(0.005)
Earliness	-0.023****	(0.001)	-0.019****	(0.003)
_cons	1.592****	(0.028)	2.850****	(0.061)
F	180.658****	χ^2	171.062****	
R^2	0.005	Pseudo r2	0.005	
Adjusted R^2	0.005			
N	1,394,399		1,394,399	
rmse	0.379			

Note: Robust standard errors reported in parenthesis. $^{+}$ $p < 0.1$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.005$, **** $p < 0.001$

Moderation effects of delivery window length and expedited

H3a predicted that delivery window length would enhance the negative effect of lateness on customer outcomes, whereas H3b predicted that delivery window length would mitigate the positive effect of earliness on customer outcomes. Results from Table 21 Models 1 and 3 provide significant coefficients for the interactions on customer rating. The plots in Figure 21 and Figure 20 suggest that a longer delivery window mitigates the effect of lateness on rating. Specifically, at higher levels of lateness, the difference between a short and long delivery window on customer rating is stronger at low levels of lateness. Hence, we do not find support for H3a. Similarly, delivery window length mitigates the effect of earliness, thus supporting H3b. The interaction coefficients on repurchase behaviors are not significant. Hence, H3a and H3b are not supported.

Table 21 – IV/2SLS second stage. Moderation effects of delivery window length and expedited delivery

	(1)		(2)		(3)		(4)	
	Ln(rating)		Ln(rating)		Repurchase		Repurchase	
order size	0.000	(0.000)	0.000****	(0.000)	0.008****	(0.000)	0.009****	(0.000)
distance from store	-0.011****	(0.002)	-0.002****	(0.000)	0.014*	(0.006)	0.017****	(0.001)
n late previous orders	-0.012****	(0.002)	-0.001*	(0.000)	0.004	(0.006)	0.011****	(0.001)
n early previous orders	-0.015****	(0.004)	0.002**	(0.001)	-0.008	(0.012)	0.008****	(0.002)
subscription	0.025****	(0.001)	0.024****	(0.001)	0.257****	(0.004)	0.255****	(0.004)
unattended	0.002*	(0.001)	0.000	(0.001)	0.034****	(0.003)	0.033****	(0.003)
Month fe	YES		YES		YES		YES	
Day of the week fe	YES		YES		YES		YES	
Clock hour fe	YES		YES		YES		YES	
Store fe	YES		YES		YES		YES	
Lateness	-0.117****	(0.018)	-0.042****	(0.001)	-0.147***	(0.052)	-0.089****	(0.005)
Earliness	-0.078****	(0.012)	-0.025****	(0.001)	-0.042	(0.033)	-0.021****	(0.003)
Window length	-0.286****	(0.042)	-0.478****	(0.056)	-0.634****	(0.115)	-0.809****	(0.145)
Expedited	-0.065****	(0.007)	0.212****	(0.036)	0.031	(0.020)	0.232*	(0.094)
Lateness x Window length	0.022****	(0.005)			0.019	(0.016)		
Earliness x Window length	0.019****	(0.004)			0.007	(0.012)		
Lateness x Expedited			-0.090****	(0.013)			-0.049	(0.036)
Earliness x Expedited			-0.087****	(0.014)			-0.076*	(0.034)
_cons	1.696****	(0.040)	2.088****	(0.065)	3.307****	(0.103)	3.684****	(0.164)
F	162.064****		162.063****	χ^2	77850.711		77854.631	
R^2	0.005		0.005	Pseudo r2	0.092		0.092	
Adjusted R^2	0.005		0.005					
N	1,394,399		1,394,399		1,394,399		1,394,399	
rmse	0.379		0.379					

Note: Robust standard errors reported in parenthesis. $^{+}$ $p < 0.1$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.005$, **** $p < 0.001$

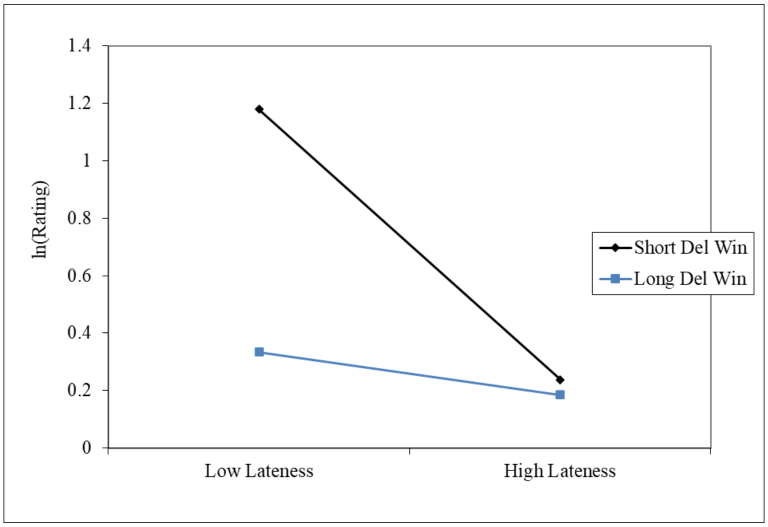


Figure 19 – Interaction effect of lateness and delivery window length on customer rating

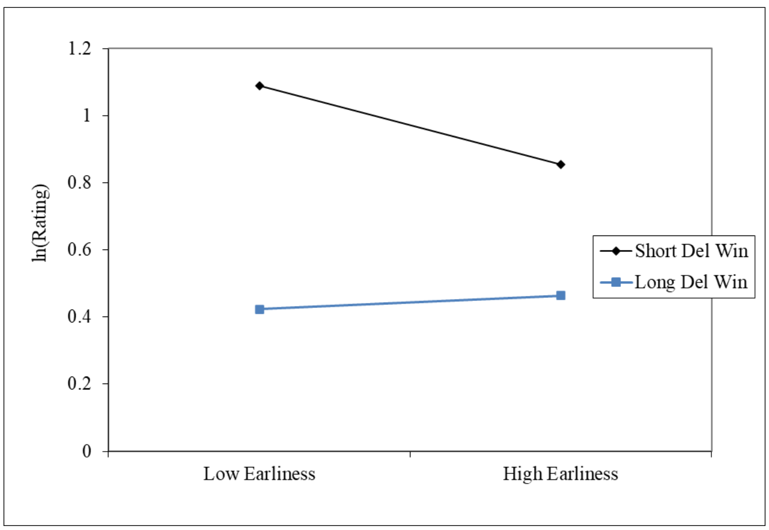


Figure 20 – Interaction effect of earliness and delivery window length on customer rating

Finally, H4a predicted that expedited delivery would enhance the negative effect of lateness on customer outcomes, whereas H4b predicted that expedited delivery would mitigate the positive effect of earliness on customer outcomes. Results from Table 21 Models 2 and 4 provide significant coefficients for the interactions on customer rating. The plots in Figure 21 and Figure 22 suggest that an expedited delivery magnifies the effect of lateness and earliness on customer rating. Hence, we find support for H4a, but not for H4b relative to customer rating. The interaction between lateness and expedited delivery presents a non-significant coefficient on repurchase

behavior, hence H4a is not supported. The interaction with earliness presents a significant coefficient. Figure 23 shows that expedited delivery enhances the effect of earliness on repurchase behavior. Thus, we do not find support for H4b.

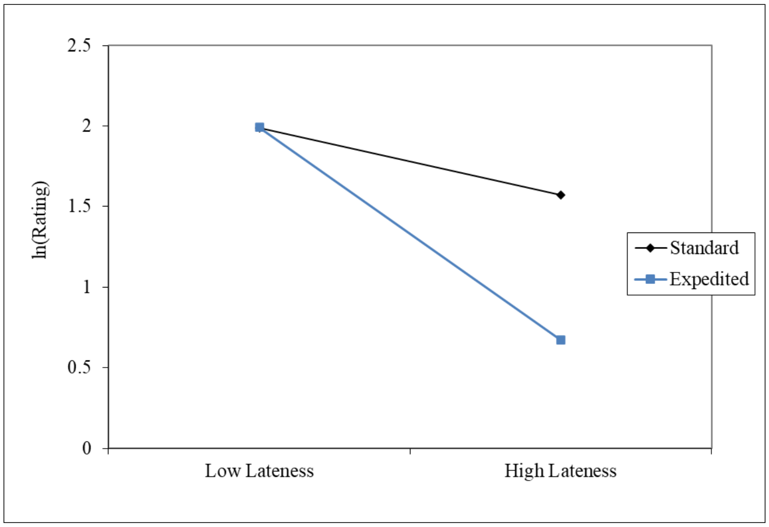


Figure 21 – Interaction effect of lateness and expedited delivery on customer rating

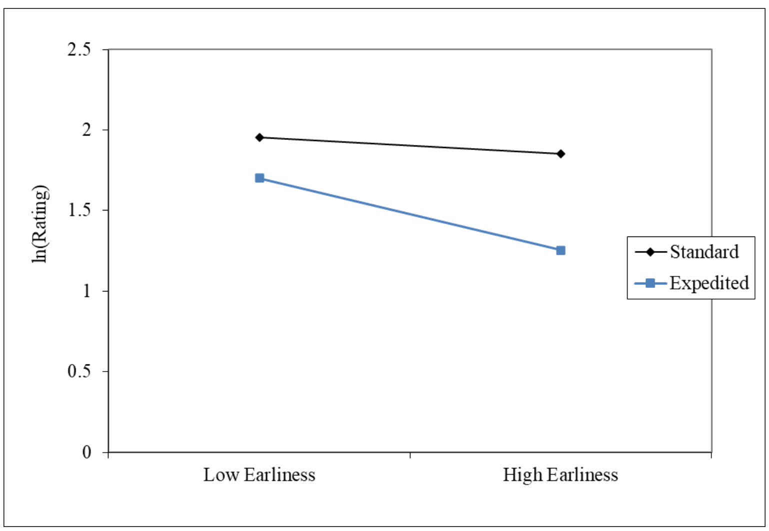


Figure 22 – Interaction effect of earliness and expedited delivery on customer rating

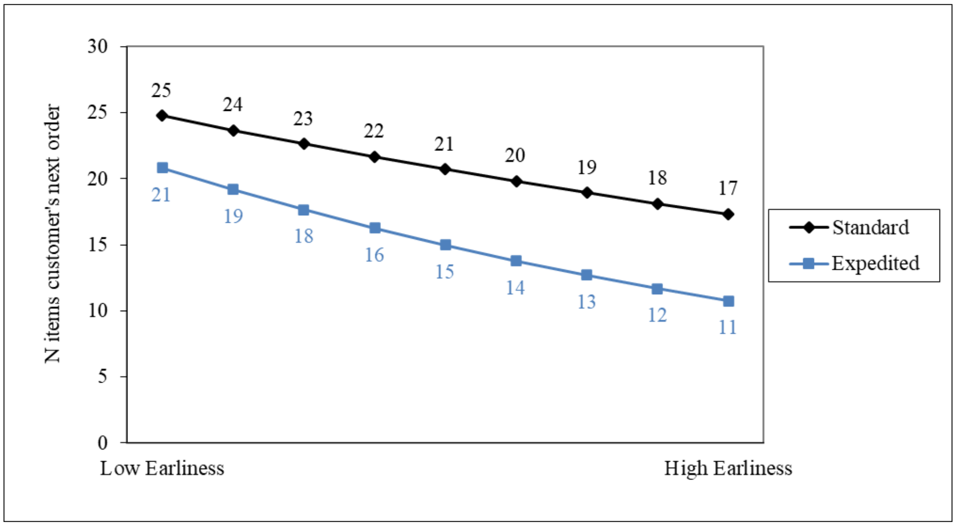


Figure 23 – Interaction effect of earliness and expedited delivery on repurchase behaviors

Post-hoc analyses

To further explore the impact of lateness and earliness on repurchase behavior, we performed two post-hoc analyses. First, we investigated the quadratic effects of lateness and earliness on repurchase behavior. Results from an OLS regression including the same set of control variables of in the main analysis show that while the quadratic effect of lateness is not significant, earliness presents a negative and significant linear coefficient ($\beta = -0.007$, SE = 0.001) but a positive and significant quadratic coefficient ($\beta = 0.0002$, SE = 0.0001). Figure 24 presents the quadratic effect of earliness on repurchase behavior following a U-shaped effect. Following recommended practices (Miller et al., 2013), we plotted the 95% lower and upper Johnson-Neyman thresholds, identifying the non-significance area between 12.2 and 17 minutes.

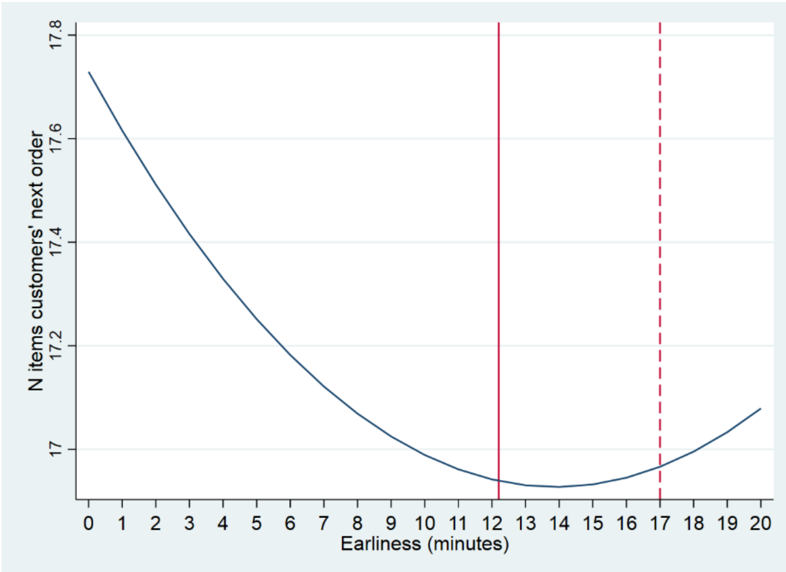


Figure 24 – Quadratic effect of earliness on repurchase behaviors with Johnson-Neyman lower 95% threshold (full red line), and upper 95% threshold (dashed red line)

Second, we performed a survival analysis that estimates the odds that a customer exits (i.e., that a customer disappears from the dataset). Given the nature of the customer exit operationalization and the characteristics of the dataset, following prior literature (Azadegan et al., 2013; De Vries et al., 2016; Singh et al., 2021), we investigate customer exit using survival analysis. Survival analysis is adopted when considering the occurrence of a binary outcome variable (presence or absence of an individual) given the left censored (late entry) and/or the right censored (truncation of the study period) nature of the dataset (Bhattacharjee et al., 2007; Flynn, 2012; Singh et al., 2021). Survival analysis can determine the impact of covariates on the survival time, defined as the time interval between the start and follow-up for a subject until the event of interest occurs or until censored (Flynn, 2012).

As common in survival analysis (Clark et al., 2003), we assess the survival rate of customers by consulting the Kaplan-Meier survival curve (Kaplan & Meier, 1958), which shows the survival function against time estimated as the cumulative probability of survival for all individuals in the dataset since the baseline (Kaplan & Meier, 1958). Thus, we identified the

cumulative probability of how many customers *survived* for how many days in the dataset before disappearing. Figure 25 presents the Kaplan-Meier curve showing that only 25% of customers (~175k for the subsample of ~702k customers) in the dataset *survive* for more than 5 days. This *does not* reflect customers that consistently purchase everyday for 5 consecutive days. Rather, it shows that only 25% of customers were still present in the dataset after 5 days since their first appearance. In addition, the curve shosws that none of the customers survived for longer than 53 days.

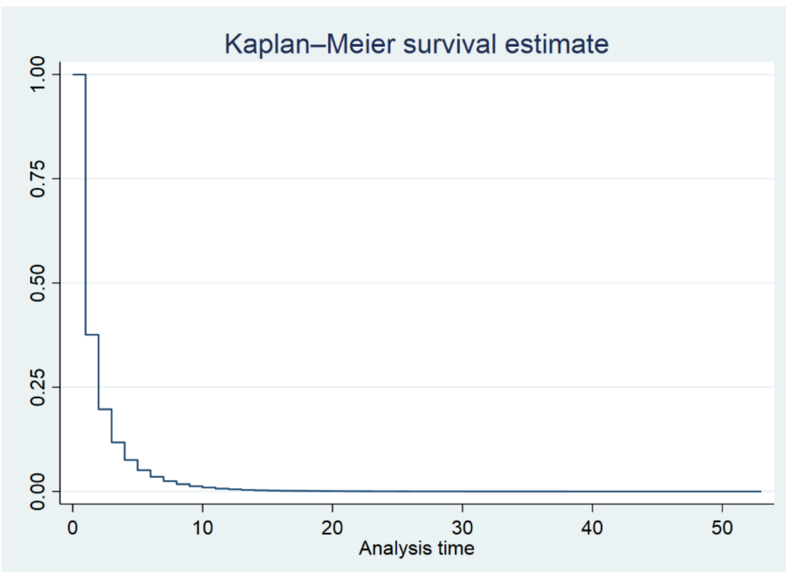


Figure 25 – Kaplan-Meier curve for customers survival analysis

Motivated by this result, we assessed the impact of lateness and earlinesson the survival rate of customers. Following prior literature (De Vries et al., 2018; Senot, 2019), we adopted a Cox proportional hazards regression model (Cox, 1972). This is a semi-parametric multivariate survival analysis model, that estimates the effect of covariates on the hazard rate, without being subject to assumptions relative to the distribution of the hazard rate over time (Bradburn et al., 2003; Senot, 2019; Singh et al., 2021). Model 1 and Model 2 in Table 22 report the estimations

for the impact of lateness and earliness on the hazard rate with control variables (equation 10) and adding the inverse mills ratio (equation 11), by assessing the following estimation models:

$$(24)\ h_d(t/\mathrm{X}) = \lambda_0(\mathrm{t}) \times \exp\{\alpha_1 \widehat{Lateness}_i + \alpha_2 \widehat{Earliness}_i + \mathrm{BX}_i\}$$

$$(25)\ h_d(t/\mathrm{X}) = \lambda_0(\mathrm{t}) \times \exp\{\alpha_1 \widehat{Lateness}_i + \alpha_2 \widehat{Earliness}_i + \mathrm{BX}_i + \alpha_3 \mathrm{imr}\}$$

Interestingly, results show that while lateness increases the odds of exit, earliness decreases the likelihood that a customer leaves the dataset. Hazard ratios below (above) 1 indicate a decrease (increase) in the chances of exit. Model 1 indicates that for 1-minute increase in lateness, holding all other variables constant, the rate of exit increases by 2.4%. Following this model, a delivery occurring after 41 minutes would result in a 100% of chance of exit. Conversely, for 1 minute increase in earliness, holding all other variables constant decreases the chances of exit by 2.3% (1.000 – 0.977 = 0.023). Similarly, a delivery completed 20 minutes earlier (the max of earliness for this dataset) would result in a 46% decrease in chances of exit.

Table 22 – Cox proportional hazard regression model for lateness and earliness

	(1) Exit β	Hazard ratio	se	(2) Exit β	Hazard ratio	se
Order size	-0.000****	1.000****	(0.000)	-0.000****	1.000****	(0.000)
Distance from store	-0.007****	0.993****	(0.001)	-0.007****	0.993****	(0.001)
N late previous orders	0.008****	1.008****	(0.001)	0.005****	1.005****	(0.001)
N early previous orders	0.007****	1.007****	(0.001)	-0.000	1.000	(0.001)
Subscription	-0.225****	0.798****	(0.002)	-0.228****	0.796****	(0.002)
Unattended	0.024****	1.024****	(0.002)	0.023****	1.023****	(0.002)
Imr				0.114****	1.121****	(0.003)
Month fe	YES			YES		
Day of the week fe	YES			YES		
Clock hour fe	YES			YES		
Store fe	YES			YES		
Lateness	0.024****	1.024****	(0.003)	0.025****	1.026****	(0.003)
Earliness	-0.023****	0.977****	(0.002)	-0.023****	0.977****	(0.002)
χ^2	180,998.618			182,557.801		
McFadden's pseudo R^2	0.006			0.006		
N	1,355,436			1,355,436		

Note: robust standard errors reported in parenthesis. $^{+}$ $p < 0.1$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.005$, **** $p < 0.001$

Study 1 – Discussion of results

Findings from Study 1 show the importance of precision in last mile delivery. In contrast to prior literature findings that distinguish between the negative effect of lateness from the positive effect of earliness on customer outcomes (Rao, Goldsby, et al., 2011; Akturk et al., 2022), the results of this study align with recent theorizations on new consumers' preferences toward the delivery service (Amorim & DeHoratius, 2021). Specifically, despite theorizing earlines as a positive disconfirmation, we found that in terms of customers' outcomes, earliness is also perceived a poor delivery performance and detrimental for customer outcomes, suggesting that delivery precision is preferred over speed (Amorim & DeHoratius, 2021; Ta et al., 2023; Zhan et al., 2023). In addition, while (Amorim & DeHoratius, 2021) report a preference for punctual deliveries (i.e., precision) for attended home deliveries, we find the resuts to be significant even controlling for the variance explained by attended vs unattended deliveries.

The moderation effects of window length and expedited delivery present differential effecs based on whether the moderator impacts both expectations and performance (window length), or just expectations (expedited delivery). The moderation effect of delivery window length intervening before purchase through forming expectations and during the delivery service through the service experience, opens the discussion to a theoretical and managerial service provision trade-off. Results reveal that, on the one hand, a shorter delivery window would increase customer expectations relative to the service delivery. Such higher expectations are harder to meet, and magnify the negative effect of service failures (lateness or earliness) on customer outcomes. On the other hand, a longer delivery window reduces customer outcomes, especially at low levels of lateness and earliness. Customers typically find unpleasant to wait longer to receive the order, as it increases uncertainty on the wait time (Donohue et al., 2020), though this depends on their

perception of waiting time (Zohar et al., 2002) (In study 2 we control for respondents' time orientation). However, in light of the fact in Figure 19 and Figure 20, short delivery windows present higher levels of customer outcomes, results would suggest that enhancing the experience is prioritized over managing expectations. Finally, the non-significant results relative to the moderation effect on repurchase behaviors suggest that customers, on their next repurchase, may develop *new* expectations depending on the window length of the next order delivery. Interestingly, expectations on the delivery window do not have a carry-over effect on the customer evaluation process of the third and fourth steps of EDT (Oliver, 1980; Bhattacherjee, 2001).

In contrast, the retailer's promise for a shorter lead time through an expedited delivery affects only prior expectations. Results showing a magnifying effect of expedited delivery at increasing levels of the service failure align with the notion that higher expectations are harder to meet, and more detrimental in the presence of a negative disconfirmation. Customers paying an extra fee for a faster delivery should result in extra-careful consideration for such orders. Interestingly, results show a carryover effect of expedited delivery on the next purchase when the delivery was early, the same does not hold when the delivery is late. As we discuss in the next paragraph, we argue that such effect does not apply to the next order size for late deliveries because customers are less likely to repurchase again. In contrast, this carry-over effect influences the next purchase after experiencing an early delivery as customers, while placing the next order, will recall the negative service experience.

Finally, the effect of delivery performance on repurchase behaviors also shows contrasting results. On the one hand, earliness and lateness reduce the customer's spending on the next order. On the other hand, the quadratic effect of earliness on repurchase behaviors and the survival analysis informs that while customer are less likely to ever repurchase after a late delivery, they

will return after an early delivery. For the latter, we conclude that customers may be willing to repurchase from the retailer, but would significantly reduce their spending. As suggested in recent literature (Lee et al., 2015; Peinkofer & Jin, 2022), an explanation could reside in how service performance and prior expectations may affect/erode levels of trust and commitment in the customer-retailer relationship. In the next study, we explore and test hypotheses based on Commitment Trust Theory (Morgan & Hunt, 1994).

Commitment Trust Theory

Commitment Trust Theory (CTT) derives from the relationship marketing framework, and suggests that relational exchanges between partners determine the success and failure of "establishing, developing, and maintaining successful relational exchanges" (Morgan & Hunt, 1994, p. 22). CTT specifies that consumers fall within the category of exchanging partners (Morgan & Hunt, 1994; Umashankar et al., 2017). Indeed, often customers have been seen as an exchanging partner trusting and committing effort to the relationship with the retailer (Garbarino & Johnson, 1999; Sirdeshmukh et al., 2002; Davis-Sramek et al., 2009; Umashankar et al., 2017). CTT identifies two key determinants of relationship efforts, namely trust and commitment (Kwon & Suh, 2004). Trust is defined as "confidence in an exchange partner's reliability and integrity" (Morgan & Hunt, 1994, p. 23). Relationship commitment is defined as the exchange partner's belief on the importance of the relationship and effort to work on the relationship to ensure its continuity (Morgan & Hunt, 1994; Umashankar et al., 2017). The role of trust and commitment has been investigated in prior operations and service management literature to determine the value generated for final customers through customer outcomes (Lee et al., 2015; Dobrzykowski et al., 2020; Jin et al., 2022; Peinkofer & Jin, 2022).

In the weakly tied retailer-customer relationship, upon experiencing a service failure, customers are likely feeling a trust breach in the reliable exchange part (Umashankar et al., 2017). Since trust is among the most important factors in the relationship, an erosion of trust will detrimentally affect customer outcomes (Peinkofer & Jin, 2022). Differently, a service failure affects commitment because in a weakly tied relationship, the exchanging partner (i.e., the customer) would not provide extra effort to maintain the relationship and will reduce customer outcomes (Umashankar et al., 2017). However, continuous interaction between these two parties will provide organizations with useful information about customers' behavior and help organizations satisfy customers, which ultimately impacts business performance (Lee et al., 2015). Thus, we hypothesize the following mediation hypotheses:

Hypothesis H5a: Trust will partially mediate the negative relationship between delivery performance and customer outcomes.

Hypothesis H5b: Commitment will partially mediate the negative relationship between delivery performance and customer outcomes.

CTT suggests that behavioral uncertainty, defined as inability to predict a partners' behavior or changes in the external environment (Umashankar et al., 2017), decreases trust of its trading partner since it creates a performance evaluation problem. A longer delivery window will increase the uncertainty of when the delivery is performed. Hence, it will magnify the impact of a service failure on complaining behaviors.

Hypothesis H6a: Delivery window moderates the indirect effect of delivery performance on customer outcomes via trust. Specifically, for a longer delivery window, a stronger negative effect is expected.

Hypothesis H6b Delivery window moderates the indirect effect of delivery performance on customer outcomes via trust. Specifically, for a longer delivery window, a stronger negative effect is expected.

Study 2 – Experiment

We investigated this second set of hypotheses using a scenario-based experiment, which constitute an appropriate approach to investigate the impact of retailer strategies on customer outcomes (Eckerd et al., 2021).

Experimental design and vignette creation

The experimental design was based on the results of Study 1 and prior operations management literature. The results of Study 1 suggests three levels of delivery performance, namely on-time, late, and early. Prior literature adopting scenario-based experiments operationalized delivery window length using three increasing levels of length (Agatz et al., 2021). In our scenario, the three levels are two, five, and eight hours. Hence, we employed a 3 (delivery performance: on-time, late, early) x 3 (window length: two, five, and eight hours) between subject scenario-based experiment.

We developed the vignette for the pretest and the main study following recommendations from (Rungtusanatham et al., 2011). Specifically, in the pre-design stage, we carefully review extant literature by conducting similar experiments to gain insights relative to the vignette length, language, and context description (Agatz et al., 2021; Thomas et al., 2022). Next, we design the common and experimental modules of the scenario as an online grocery shopping order placement (see scenarios in Appendix D). In the common module, the vignette specified the purchase of a week's worth of grocery delivery at Grocer.com, a fictitious online grocery retailer providing the delivery service. Fictitious names are typically used in scenario-based experiments to control for

brand effect and avoid that familiarity could influence the responses (Abdulla et al., 2022; Mollenkopf et al., 2022; Peinkofer & Jin, 2022). We chose a week's worth of groceries to align with prior literature and ensure generalizability. Specifically, Agatz et al. (2021) offered participants a chance to win $100-voucher to shop groceries. Recent evidence found that US customers spend, on average, $5,259/year (~$100/week) for food at home (U.S. Bureau of Labor Statistics, 2022). As confirmation, we asked participants to indicate their average weekly bill for groceries, and found consistent estimations in the pretest (M = $173.11, SD = 105.61) and in the main study (M = $154.63, SD = 98.01). Table 8 presents the demographics of the samples used for the pretest and the main study.

In the experimental module, which varied across experimental conditions (Rungtusanatham et al., 2011), respondents were randomly assigned to one of the three treatments of the delivery window. This treatment informed participants that they chose a specific delivery window length because of most convenient for them. Then, within each window length treatment, we randomly assigned respondents to one alternative time slots of the same time length. Hence, Participants were asked to select the time slot. In total, we counted five time window alternatives for the 2 hours delivery window (10 AM – 12 PM; 12 PM – 2 PM; 2 PM – 4 PM; 4 PM – 6 PM; 6 PM – 8 PM), three alternatives for the 5 hours delivery window (10 AM – 3 PM; 1 PM – 6 PM; 3 PM – 8 PM), and two alternatives for the 2 hours delivery window (10 AM – 6 PM; 12 PM – 8 PM). This procedure follows (Agatz et al., 2021), who advocates for this experimental design as it avoids sampling bias. Evidence provided in the section Experimental checks suggests that this design did not affect how respondents perceived the window length.

Next, participants were exposed to the treatment relative to the delivery performance. The vignette showed the screenshow of a completed delivery email informing respondents on the

delivery window length and delivery time (see Appendix D). Recent industry reports and literature suggest that emails are the common communication tool to provide consumers with delivery updates (Retail TouchPoints et al., 2018; Russo et al., 2022). For the on-time delivery, the manipulation was consistent across all delivery windows and informed participants that the delivery occurred 15 minutes passed the beginning of the delivery window (i.e., still within the delivery window). For major delivery platforms, 15-20 minutes is a threshold to account for an on-time delivery when evaluating a driver's performance (Doordash.com, 2023; Walmart.com, 2023). This time cushion excludes that the performance of a driver would be affected by external factors, such as wait time at the merchant. For late and early delivery, we chose to expose customers to one-hour late and one-hour early delivery, respectively. Holding the experimental characteristics constant across treatments ensured to control for potential demand effect rising from different levels of lateness and earliness (Lonati et al., 2018). After the vignette, participants were asked to complete a survey and respond to demographic questions (see Table 23).

Participants

We recruited participants for the pretest and main study from Prolific Academic, an online platform often used in operations management literature adopting scenario-based experiments (Bhatia, 2019; Schneider et al., 2021; Serra-Garcia & Szech, 2023). To ensure the quality of responses, we recruited participants meeting a set of specific criteria, namely: Reside in the United States, be at least 18 years old, have at least a 90% approval rate for previously completed studies, and, for the main study have not participated to the pretest. Upon completing the task (~ 10 minutes), each participants received $2. We held the compensation constant for both the pretest and the main study. Finally, we used a Qualtrics randomizer to randomly assign respondents to the nine treatments. We collected a total of 111 responses for the pretest and 642 for the main study

(see Table 23). Upon removing participants who failed the attention check, manipulation checks, and did not complete the task, the final sample was 81 for the pretest, and 472 for the main study. For the main study, we met the threshold 50 observations per cell (Lonati et al., 2018).

Table 23 – Summary of participants demographisc and experimental procedures

	Pretest	**Main study**
Initial sample	111	640
Failed screening questions	9	51
Failed attention check	10	17
Incomplete survey	2	23
Failed memory recall check	9	77
Final sample size	81	472
Compensation	$2.00	$2.00
Average completion time	9 min 25 s	10 min 8 s
Average age	34.79	38.12
% Female	43%	49%
Percentage with some college	72.22%	71.19%
Median household income	$50,000 - $74,999	$50,000 - $74,999
Purchase channel (% online)	48.89%	47.46%
Delivery area (% Urban)	67%	70%
Married (% Married)	54%	54.89%
Child (% No child)	57%	60.21%
Employed (% Full time)	42%	51.49%
Fam size (% 1 or 2 members)	41%	52.55%
Home (% own)	42%	55.96%
Avg weekly grocery bill	$ 173.11	$ 154.63
Ethnicity (%White)	60%	56.75%

Experimental checks

Following common practice (Bachrach & Bendoly, 2011; Lonati et al., 2018), we conducted a series of experimental checks to ensure the validity of our manipulations. Table 24 presents the results of the experimental checks. Below a detailed description of the results.

Table 24 – Experimental checks

		Delivery performance			Window length		
		OT vs L	**OT vs E**	**E vs L**	**2h vs 5h**	**2h vs 8h**	**5h vs 8h**
Man check 7-point scale	P	$F_{1,50} = 125.15^*$	$F_{1,53} = 123.83^*$	$F_{1,53} = 508.13^*$	$F_{1,54} = 290.81^*$	$F_{1,52} = 331.72^*$	$F_{1,50} = 75.64^*$
	MS	$F_{1,313} = 974.23^*$	$F_{1,315} = 877.54^*$	$F_{1,310} = 2{,}977.05^*$	$F_{1,319} = 5{,}672.75^*$	$F_{1,308} = 3{,}893.12^*$	$F_{1,311} = 905.41^*$
Man check (% correct)	P	96.30%			93.83%		
	MS	97.25%			96.82%		
Confounding check	P	$F_{2,78} = 0.01$ $p = 0.99$			$F_{2,78} = 1.05$ $p = 0.35$		
	MS	$F_{2,469} = 0.10$ $p = 0.90$			$F_{2,469} = 0.27$ $p = 0.76$		
Hawthorne 1	P	$F_{2,78} = 0.05$ $p = 0.96$			$F_{2,78} = 0.08$ $p = 0.93$		
	MS	$F_{2,469} = 0.62$ $p = 0.54$			$F_{2,469} = 0.32$ $p = 0.72$		
Hawthorne 2	P	$F_{2,78} = 0.03$ $p = 0.98$			$F_{2,78} = 0.98$ $p = 0.38$		
	MS	$F_{2,469} = 1.11$ $p = 0.33$			$F_{2,469} = 1.30$ $p = 0.28$		
Hawthorne 3	P	$F_{2,78} = 1.93$ $p = 0.15$			$F_{2,78} = 0.68$ $p = 0.51$		
	MS	$F_{2,469} = 0.75$ $p = 0.48$			$F_{2,469} = 1.10$ $p = 0.33$		
Hypotheses awareness	MS	$F_{1,313} = 0.04$ $p = 0.84$	$F_{1,315} = 0.39$ $p = 0.53$	$F_{1,310} = 0.67$ $p = 0.42$	$F_{1,319} = 1.94$ $p = 0.17$	$F_{1,308} = 1.00$ $p = 0.32$	$F_{1,311} = 0.09$ $p = 0.76$
Realism 1	P	M = 5.99 SD = 1.26					
	MS	M = 5.93 SD = 1.32					
Realism 2	P	M = 6.04 SD = 0.95					
	MS	M = 6.03 SD = 0.94					
Time of the day effect					**2h**	**5h**	**8h**
					$F_{4,24} = 0.75$ $p = 0.57$	$F_{2,24} = 0.13$ $p = 0.88$	$F_{1,24} = 4.09$ $p = 0.06$

Note: * $p < 0.01$; OT = On-time; L = Late; E = Early; 2h = Two hours; 5h = Five hours; 8h = Eight hours; P = Pretest; MS = Main Study

First, we conducted a manipulation check in the form of a memory recall check (Abbey & Meloy, 2017), asking participants to rate on a 7-point scale (1) when the order arrived (1 = Very early; 4 = On Time; 7 = Very late), and (2) the length of the delivery window (1 = Two hours; 4 = 5 hours; 7 = Eight hours). Table 24 reports the results from a one-way ANOVA comparing each pair of treatments. From the initial sample, we dropped participants who failed this manipulation check. Specifically, those exposed to the early condition that rated the order as being delivered on-time or any level of late (N = 13); those exposed to the on-time condition that rated the order delivered very early or very late (N = 60); and those exposed to the late condition that rate the order delivered on-time or any level of early (N = 4). In a post-hoc analysis, we performed the main analysis, including these removed observations, and found consistent estimations. Next, we confirmed that no confounding effects across treatments occurred through a one-way ANOVA. In

addition, we ensured that the different timing of delivery window options did not significantly affect participants' response to this manipulation check. In the pretest, within group one-way ANOVA revealed no significant differences among delivery window options. Finally, we also conducted an additional manipulation check in the form of a memory recall check (Abbey & Meloy, 2017), asking participants to recall (1) when the order arrived (Late/Early/On-time), and (2) the length of the selected delivery window (Two hours/Five hours/Eight hours) (Ma, 2017). We found that both in the pretest and main study, the percentages of respondents who correctly guessed the manipulations were all above 90%.

We conducted Hawthorne checks to assess if treatments could change the participants goals and motivations. Three items asking participants to rate the importance of (1) ensuring that the product is traceable during the delivery, (2) the price of the product, (3) the price of the subscription plan. We found no significant differences across groups, suggesting that the experimental conditions did not influence the participants’ three goals.

We ensured the quality of responses by including an attention check in the form of a direct query (Abbey & Meloy, 2017), screening questions, and several timing checks capturing how long respondents spent to complete the task, meeting our 10-minute task expectations for the pretest ($M = 9.41$, $SD = 4.87$) and the main study ($M = 10.14$, $SD = 6.52$). Realism checks were captured on both the pretest and main study on a 7-point Likert scale adapted from (Dabholkar et al., 1996), asking participants to rate the extent to which (1) the difficulty to imagine myself in the shopping situation, and (2) the described shopping scenario was realistic. Results report means ranging from 5.93 to 6.07, suggesting the scenario was deemed realistic.

Finally, we verified whether participants were aware of the research hypotheses, capturing Perceived awareness of research hypotheses on a 4-item 7-point Likert scale adapted from ($\alpha =$

0.95) (Rubin et al., 2010; Nichols et al., 2019). The scale asked participants to rate the extent to which they (1) know what the researchers are trying to discover in this research, (2) do not know what the researchers are trying to prove in this research (r), (3) have a good idea about what the predictions are in this study, and (4) are not sure exactly what the researchers are aiming to prove in this research (r). One sample t-test shows that the scale mean ($M = 3.47$, $SD = 1.41$) is significantly lower than the scale's midpoint of 4.00, $t(384) = -7.36$, $p < 0.01$. In addition, ANOVA comparisons produce non-significant results (Table 24). Hence, none of the treatments significantly improved participants' guessing of the research hypotheses.

Measures

While the independent variable and moderator of this study were manipulated, the mediators, dependent variables, and one covariate were captured using multi-item behavioral scales. All scales were captured on a 7-point Likert scale. We captured Trust and Commitment with two five-item scales adapted from (Morgan & Hunt, 1994). Satisfaction with the retailer was captured with a four-item scale adapted from (Garbarino & Johnson, 1999; Davis-Sramek et al., 2009), and repurchase behaviors with a four-item scale adapted from (Zeithaml et al., 1996). We capture time orientation, a control variable included in the main estimations, with a four-item scale adapted from (Kaufman et al., 1991). Table 25 reports each scale items, and Table 26 the pairwise correlations. Finally, we captured several respondents' demographics that we operationalized for the main analysis, as reported in Table 27.

Study 2 – Results

Measurement model

We conducted a confirmatory factor analysis (CFA) on AMOS 28 of a five-factor model, including trust, commitment, satisfaction with the retailer, repurchase intentions, and time orientation, using

maximum likelihood estimation. Table 25 summarizes factor loadings, descriptive statistics, and model fit statistics. The CFA indicates that the model fit statistics are within acceptable limits (Hu & Bentler, 1999). Convergent validity was confirmed by the average variance extracted (AVE) and the factor loadings for all substantive variables exceeding the recommended threshold of 0.50 (Fornell & Larcker, 1981). Discriminant validity was confirmed since each pair of constructs' squared correlation was less than the AVE for each pair of variables (Fornell & Larcker, 1981). The Cronbach's alpha(s) were all above the 0.70 threshold, indicating internal consistency (reliability) of the scale (Nunnally & Bernstein, 1994). Finally, composite reliability (CR) values exceeded the minimum threshold of 0.60 (Bagozzi & Yi, 1988).

Table 25 – Confirmatory factor analysis and scale items statistics

Measure and individual items	
Satisfaction with the retailer (7-point Likert scale: 1= Strongly disagree, 7 = Strongly agree), adapted from (Garbarino & Johnson, 1999; Davis-Sramek et al., 2009)	M = 4.08, SD = 1.76
"Based on the scenario, to what extent do you agree with the following statements?"	
Grocer.com comes close to giving me the "perfect" shopping experience	0.943
My current shopping experience with Grocer.com has been superior	0.966
Grocer.com sets itself apart from others because of its superior service	0.933
Grocer.com sets itself apart from others because of its superior service	0.924
AVE = 0.89; Cronbach = 0.97, CR = 0.97	
Repurchase intentions (7-point Likert scale: 1= Strongly disagree, 7 = Strongly agree), adapted from (Zeithaml et al., 1996)	M = 4.21, SD = 1.80
"Based on the scenario, to what extent do you agree with the following statements?"	
I would continue shopping from Grocer.com	0.956
The next time I need groceries, I will purchase from Grocer.com	0.973
I would consider Grocer.com my first choice to grocery shopping	0.943
I will use the service of Grocer.com in the next months	0.941
AVE = 0.91; Cronbach = 0.98, CR = 0.98	
Trust (7-point Likert scale: 1= Strongly disagree, 7 = Strongly agree), adapted from (Morgan & Hunt, 1994)	M = 4.18,
"Based on the scenario, to what extent do you agree with the following statements?"	SD = 1.55
I feel that I can trust Grocer.com completely	0.823
Grocer.com is truly sincere in its promises	0.874
Grocer.com is honest and truthful with me	0.868
Grocer.com treats me fairly and justly	0.805
I feel that Grocer.com can be counted on to help me when I need it	0.838
AVE = 0.71; Cronbach = 0.92, CR = 0.92	

Table 25 (Cont.)

Commitment (7-point Likert scale: 1= Strongly disagree, 7 = Strongly agree), adapted from (Morgan & Hunt, 1994)	M = 3.71, SD = 1.55
"Based on the scenario, to what extent do you agree with the following statements?"	
I am proud to be a customer of Grocer.com	0.856
I feel a sense of connection to Grocer.com	0.78
I care about the long-term success of Grocer.com	0.783
I am a loyal patron of Grocer.com	0.881
I plan to maintain a long-term relationship with Grocer.com	0.893
AVE = 0.71; Cronbach = 0.92, CR = 0.92	
Time orientation (7-point Likert scale: 1= Strongly disagree, 7 = Strongly agree), adapted from (Kaufman et al., 1991)	M = 4.09, SD = 0.62
Based on your personal preferences, to what extent do you agree with the following statements?	
I do not like to juggle several activities at the same time	0.853
People should not try to do many things at once	0.584
When I sit down at my desk, I work on one project at a time	0.617
I am comfortable doing several things at the same time (r)	0.725
AVE = 0.50; Cronbach = 0.79, CR = 0.79	
Model fit	
χ^2 = 594.07, d.f. = 284, $\chi^2/d.f.$ = 2.09 ($p < 0.01$)	
SRMR = 0.031, GFI = 0.91, CFI = 0.97, RMSEA = 0.05 (90%), CI [0.04; 0.05], IFI = 0.97, TLI = 0.97.	

Note: AVE = Average Variance Extracted. α = Cronbach's alpha. CR = Composite Reliability. SRMR = Standardized Root Mean Residual. GFI = Goodness of Fit Index. RMSEA = Root Mean Square Error of Approximation. CFI = Confirmatory Fit Index. TLI = Tucker-Lewis Index. IFI = Incremental Fit Index.

Table 26 – Pairwise correlations of substantive variables

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
(1) Satisfaction with the retailer	1.00						
(2) Satisfaction with the delivery service	0.90^{***}	1.00					
(3) Repurchase intentions	0.89^{***}	0.84^{***}	1.00				
(4) Service recovery expectations	0.54^{***}	0.51^{***}	0.52^{***}	1.00			
(5) Trust	0.68^{***}	0.68^{***}	0.67^{***}	0.43^{***}	1.00		
(6) Commitment	0.76^{***}	0.70^{***}	0.79^{***}	0.48^{***}	0.60^{***}	1.00	
(7) Time orientation	0.04	0.04	0.02	0.03	0.04	0.05	1.00

Note: * $p < 0.05$

Table 27 – Operationalization of covariates included in the mediation analysis

Covariate	**Type**	**Definition**	**Source**
Age	Ordinal	Participant's age	(Abdulla et al., 2022)
Gender	Categorical	1 = Female; 0 = Otherwise	(Abdulla et al., 2022)
Income	Binary	1 = Income above sample median; 0 = Otherwise	(Confente et al., 2021)
Education	Binary	1 = Some college degree; 0 = Otherwise	(Confente et al., 2021)
Channel	Binary	1 = Online; 0 = Otherwise	(Abdulla et al., 2022)

Common method bias

We assessed the presence of common method bias using the marker variable procedure recommended by Williams et al. (2010) and Craighead et al. (2011), and used in recent operations management literature (Arellano et al., 2021; Franke et al., 2022).

The marker variable technique requires the inclusion of a theoretically unrelated construct in the survey to determine if the CFA model improves as we include such marker variable. In this study, we use a variable purposely developed for use as a marker, blue attitude (Simmering et al., 2015). Blue attitude is a four-item scale measured on a 7-point Likert scale (1 = Strongly disagree, 7 = Strongly agree) that assesses the extent to which participants (1) prefer blue to other colors (2) like the color blue (3) like blue clothes (4) hope their next car is blue (α = 0.75).

Table 28 – Common method bias tests

Model		**χ^2 (*df*)**	**CFI**	**RMSEA (90% CI)**	**Model description**
1	CFA with marker variable	822.43 (384)	0.963	0.049 (0.045, 0.054)	The CFA model with the substantive variables and the marker variable
2	Baseline (uncorrelated marker)	825.89 (397)	0.964	0.048 (0.043, 0.052)	Model 1 but the marker variable is not allowed to correlate with the other variables, and the marker items error terms and factor loadings set to the unstandardized values from model 1
3	Method-C (constrained)	825.08 (396)	0.964	0.048 (0.043, 0.052)	Model 2 with secondary loadings equal to one another from the marker variable to the substantive variables items
4	Method-U (unconstrained)	807.93 (375)	0.963	0.050 (0.045, 0.054)	Model 2 with secondary loadings allowed to vary freely from the marker variable to the substantive variables items
χ^2 model comparison		**$\Delta\chi^2$**	**Δdf**	**χ^2 critical value**	
Baseline vs Method-C		0.80	1	3.84	
Method-C vs Method-U		17.2	21	32.67	

Notes: CFA =confirmatory factor analysis; CFI = comparative fit index; RMSEA = root mean square error of approximation.

We estimate and compare the χ^2 test for four models: CFA with marker variable, a baseline model, method-C model, and method-U model (detailed description and results in Table 28). We find that modeling the marker variable additional item loadings onto all other substantive variables does not improve the model fit. The χ^2 comparison test show insignificant results ($\Delta\chi^2 = 0.80$, Δdf

= 1, $\Delta\chi^2$critical = 3.84) Thus, common method bias does not constitute a significant bias to our estimations.

Analysis and Results

We conducted a mediation and moderated mediation analysis via bootstrapping (Rungtusanatham et al., 2014), using Process Macro v4.3 for SPSS 28 (Hayes, 2018). This estimation tool is based on ordinary least-squares regression analysis, and effectively estimates mediation effects through bootstrap confidence intervals (Abdulla et al., 2022; Peinkofer et al., 2022).

We tested the mediation effect for each relevant comparison (e.g., On-Time vs Late; On-Time vs Early; Early vs Late) using Process Model 4 for parallel mediation, with trust and commitment as mediators, and separately satisfaction with the retailer and repurchase intentions as dependent variables (see Figure 26 for the statistical representation of parallel mediation). For each relevant comparison, the independent variable was operationalized as a binary variable, for example, On-Time = 0 and Late = 1. We performed the analysis including Age, Gender, Education, Income, Channel, Time orientation as covariates, and with 5,000 bootstrap resamples for te analysis. Table 29 reports the results of the mediation analysis, including direct effects, indirect effects, and 95% confidence intervals. The reported direct and indirect effects are all significant at $p < 0.05$, since the confidence interval does not include zero (Hayes, 2018).

H5a,b predicted the effect of delivery performance on customer outcomes to be mediated via trust and commitment. We found that trust and commitment each significantly mediate the negative relationship between the focal level for every comparison and the dependent variable. The significant indirect effects suggest that a late (early) delivery, as compared to on-time delivery, significantly reduce trust, commitment, satisfaction with the retailer, and repurchase intentions.

Between an early and late delivery, late deliveries present worse results. Hence, H5a,b are supported.

Process Model 4 – Parallel Mediation

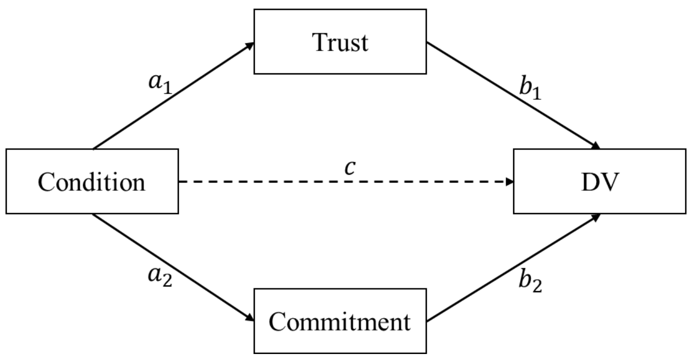


Figure 26 – Statistical diagram and notations for mediation analysis Process Model 4

Table 29 – Mediation results of Process Model 4 with parallel mediation for the effect of Late (L) and Early (E) on Satisfaction with the retailer (SAT) and Repurchase intentions (REP) via Trust (T) and Commitment (C)

Condition	**Focal lever**	**Mediation paths effects** $X \xrightarrow{a} M \xrightarrow{b} DV$	**Indirect effects [95% CI]** $\theta = a \times b$	**Direct effect [CI]** $X \xrightarrow{c} DV$
On-time vs Late	Late	$L \xrightarrow{-1.805} T \xrightarrow{0.241} SAT$	-0.436 [-0.670; -0.249]	$L \xrightarrow{-1.452} SAT$
		$L \xrightarrow{-1.861} C \xrightarrow{0.434} SAT$	-0.808 [-1.078; -0.555]	
		$L \xrightarrow{-1.805} T \xrightarrow{0.220} REP$	-0.397 [-0.624; -0.212]	$L \xrightarrow{-0.977} REP$
		$L \xrightarrow{-1.861} C \xrightarrow{0.609} REP$	-1.133 [-1.425; -0.867]	
On-time vs Early	Early	$E \xrightarrow{-1.125} T \xrightarrow{0.311} SAT$	-0.349 [-0.517; -0.212]	$E \xrightarrow{-0.721} SAT$
		$E \xrightarrow{-1.056} C \xrightarrow{0.561} SAT$	-0.592 [-0.826; -0.386]	
		$E \xrightarrow{-1.125} T \xrightarrow{0.291} REP$	-0.327 [-0.491; -0.197]	$E \xrightarrow{-0.397} REP$
		$E \xrightarrow{-1.056} C \xrightarrow{0.681} REP$	-0.728 [-0.984; -0.488]	
Early vs Late	Late	$E \xrightarrow{-0.723} T \xrightarrow{0.343} SAT$	-0.248 [-0.411; -0.115]	$L \xrightarrow{-0.316} SAT$
		$E \xrightarrow{-0.859} C \xrightarrow{0.603} SAT$	-0.518 [-0.734; -0.325]	
		$E \xrightarrow{-0.723} T \xrightarrow{0.368} REP$	-0.159 [-0.256; -0.079]	$L \xrightarrow{-0.273} REP$
		$E \xrightarrow{-0.859} C \xrightarrow{0.682} REP$	-0.350 [-0.485; -0.229]	

Note: Reported effects are all significant at $p < 0.05$. Controls included: Age, Gender, Education, Income, Channel, Time orientation. Results with 5,000 bootstraps

Next, we performed the moderated mediation analysis with moderation in the first of mediation using Process Model 7 (see Figure 27 for the statistical representation of parallel mediation). The moderator, window length was coded as multicategorical. Thus, we set the

moderator as indicator in the multicategorical option on process macro for SPSS. This option takes the first group (2 hours) as the reference group, and generates two binary variables (WL1 and WL2) that compares the reference group to the second group (5 hours), and third group (8 hours) (Hayes & Montoya, 2017). The significance of the moderated mediation depends on the significance of the Index of Moderated Mediation confidence interval (i.e., the CI does not include zero) (Hayes, 2018). Tables 30-31-32 reports the results for the first stage of mediation, the conditional indirect effects, and index of moderated mediation.

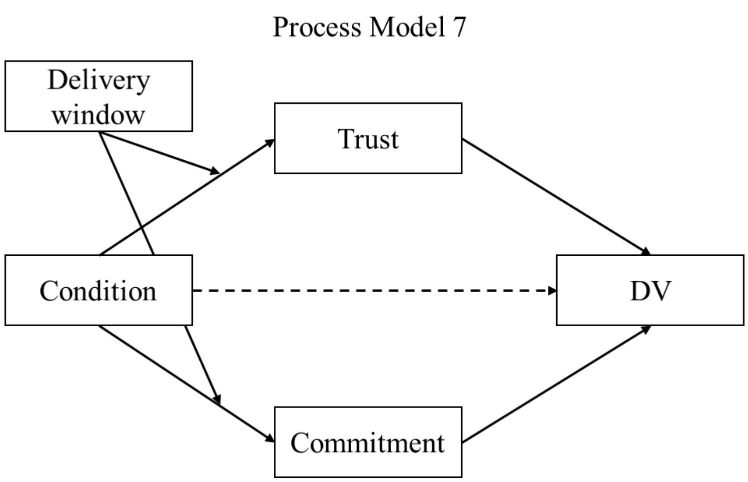


Figure 27 – Statistical diagram and notations for mediation analysis Process Model 7

Results show that window length does not moderate the indirect effects via trust. Thus, H6a is not supported. Conversely, the index of moderated mediation for the indirect effects via commitment are significant for the comparison two hours vs eight hours. Hence, H6b is partially supported. We plotted such significant interactions in Figure 28 for the On-time vs Late comparison, and Figure 29 for the comparison On-time vs Late. Overall, the plots show lower levels of commitment for late, early, and longer delivery window. However, while the effect of window length is stronger for an on-time delivery but does not affect commitment for a late delivery, in the on-time vs. early comparison, the effect of window length is stronger for an on-time delivery but does not affect commitment for late delivery.

Table 30 – Moderated Mediation results for the effect of On-time vs Late on Satisfaction with the retailer (SAT) and Repurchase Intentions (REP), via Trust (T) and Commitment (C), moderated by Window Length (WL)

On-time vs Late	(1) **Trust**	(2) **Commitment**	(3) **SAT**	(4) **REP**
Constant	4.732*** (0.582)	4.111*** (0.519)	2.27*** (0.426)	1.672*** (0.461)
Late	-1.879*** (0.25)	-2.125*** (0.223)	-1.452*** (0.135)	-0.977*** (0.146)
Trust			0.434*** (0.048)	0.609*** (0.051)
Commitment			0.241*** (0.043)	0.22*** (0.046)
WL 1 (2h vs 5h)	-0.26 (0.249)	-0.197 (0.222)		
WL 2 (2h vs 8h)	-0.424* (0.252)	-0.636*** (0.225)		
Late x WL 1	0.053 (0.351)	0.183 (0.313)		
Late x WL 2	0.189 (0.36)	0.644** (0.321)		
Age	0.006 (0.005)	0.009* (0.005)	-0.006 (0.004)	-0.003 (0.004)
Gender	0.172 (0.156)	-0.163 (0.139)	-0.006 (0.107)	-0.061 (0.116)
Education	-0.266 (0.168)	-0.061 (0.15)	-0.131 (0.114)	0.177 (0.124)
Income	-0.174 (0.145)	0.044 (0.13)	-0.158 (0.1)	-0.305*** (0.108)
Channel	0.111 (0.146)	0.272** (0.13)	0.096 (0.099)	0.164 (0.108)
Time orientation	0.124 (0.121)	0.12 (0.108)	0.08 (0.082)	-0.014 (0.089)
F-value (*df*)	15.614*** (11,300)	21.293*** (11,300)	107.88*** (9,302)	90.99*** (9,302)
R^2	0.603	0.662	0.873	0.855
IMM indirect effect via Trust		IMM WL 1	0.013 (0.084) CI [-0.159,0.177]	0.012 (0.077) CI [-0.139,0.167]
		IMM WL 2	0.046 (0.091) CI [-0.134,0.229]	0.042 (0.084) CI [-0.115,0.218]
Indirect effect via Trust		IE WL 2 hours	-0.453 (0.117) CI [-0.706,-0.248]	-0.413 (0.118) CI [-0.667,-0.209]
		IE WL 5 hours	-0.440 (0.118) CI [-0.688,-0.235]	-0.401 (0.116) CI [-0.652,-0.203]
		IE WL 8 hours	-0.408 (0.116) CI [-0.662,-0.209]	-0.371 (0.110) CI [-0.612,-0.187]
IMM indirect effect via Commitment		IMM WL 1	0.079 (0.128) CI [-0.174,0.336]	0.111 (0.184) CI [-0.253,0.474]
		IMM WL 2	0.280 (0.139) CI [0.010,0.551]	0.392 (0.197) CI [0.013,0.793]
Indirect effect via Commitment		IE WL 2 hours	-0.923 (0.148) CI [-1.230,-0.652]	-1.294 (0.173) CI [-1.646,-0.969]
		IE WL 5 hours	-0.843 (0.159) CI [-1.181,-0.551]	-1.182 (0.194) CI [-1.582,-0.822]
		IE WL 8 hours	-0.643 (0.154) CI [-0.981,-0.365]	-0.901 (0.181) CI [-1.278,-0.566]

Note: First stage common to both SAT and REP presented in models 1 and 2. * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$. Standard error in parenthesis.
Abbreviations: IMM = Index of Moderated Mediation; IE = Indirect effect; CI = Confidence Interval.

Table 31 – Moderated Mediation results for the effect of On-time vs Early on Satisfaction with the retailer (SAT) and Repurchase Intentions (REP), via Trust (T) and Commitment (C), moderated by Window Length (WL)

On-time vs Early	(1) **Trust**	(2) **Commitment**	(3) **SAT**	(4) **REP**
Constant	5.468*** (0.641)	4.27*** (0.611)	1.986*** (0.453)	1.028** (0.478)
Early	-1.444*** (0.281)	-1.675*** (0.268)	-0.721*** (0.118)	-0.358*** (0.125)
Trust			0.561*** (0.044)	0.677*** (0.047)
Commitment			0.311*** (0.043)	0.299*** (0.045)
WL 1 (2h vs 5h)	-0.272 (0.283)	-0.149 (0.27)		
WL 2 (2h vs 8h)	-0.442 (0.286)	-0.591** (0.273)		
Early x WL 1	0.248 (0.397)	0.584 (0.379)		
Early x WL 2	0.741* (0.403)	1.324*** (0.384)		
Age	0.007 (0.006)	0.008 (0.006)	-0.01** (0.004)	-0.007 (0.004)
Gender	-0.115 (0.171)	-0.004 (0.163)	-0.018 (0.113)	0.082 (0.12)
Education	-0.113 (0.183)	0.181 (0.175)	-0.108 (0.121)	0.089 (0.128)
Income	-0.024 (0.164)	0.176 (0.157)	-0.112 (0.108)	-0.061 (0.114)
Channel	0.066 (0.164)	0.410*** (0.157)	0.112 (0.109)	0.143 (0.116)
Time orientation	-0.056 (0.133)	-0.016 (0.126)	-0.053 (0.088)	-0.026 (0.093)
F-value (*df*)	4.992*** (11,305)	6.751*** (11,305)	72.126*** (9,307)	69.546*** (9,307)
R^2	0.391	0.443	0.824	0.819
IMM indirect effect via Trust		IMM WL 1	0.077 (0.124) CI [-0.164,0.331]	0.074 (0.119) CI [-0.155,0.316]
		IMM WL 2	0.230 (0.137) CI [-0.021,0.520]	0.221 (0.134) CI [-0.021,0.508]
Indirect effect via Trust		IE WL 2 hours	-0.449 (0.119) CI [-0.711,-0.24]	-0.431 (0.117) CI [-0.684,-0.231]
		IE WL 5 hours	-0.372 (0.104) CI [-0.599,-0.186]	-0.357 (0.105) CI [-0.587,-0.175]
		IE WL 8 hours	-0.219 (0.105) CI [-0.446,-0.026]	-0.210 (0.102) CI [-0.425,-0.024]
IMM indirect effect via Commitment		IMM WL 1	0.327 (0.207) CI [-0.091,0.733]	0.395 (0.248) CI [-0.091,0.874]
		IMM WL 2	0.743 (0.228) CI [0.305,1.196]	0.896 (0.273) CI [0.384,1.441]
Indirect effect via Commitment		IE WL 2 hours	-0.939 (0.173) CI [-1.298,-0.611]	-1.134 (0.197) CI [-1.532,-0.756]
		IE WL 5 hours	-0.612 (0.160) CI [-0.932,-0.317]	-0.738 (0.194) CI [-1.140,-0.383]
		IE WL 8 hours	-0.197 (0.173) CI [-0.543,0.132]	-0.237 (0.209) CI [-0.651,0.171]

Note: First stage common to both SAT and REP presented in models 1 and 2. * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$. Standard error in parenthesis.
Abbreviations: IMM = Index of Moderated Mediation; IE = Indirect effect; CI = Confidence Interval.

Table 32 – Moderated Mediation results for the effect of Early vs Late on Satisfaction with the retailer (SAT) and Repurchase Intentions (REP), via Trust (T) and Commitment (C), moderated by Window Length (WL)

Early vs Late	(1) **Trust**	(2) **Commitment**	(3) **SAT**	(4) **REP**
Constant	4.115^{***} (0.597)	2.752^{***} (0.548)	1.155^{**} (0.457)	0.823^{*} (0.442)
Late	-0.488^{*} (0.274)	-0.463^{*} (0.251)	-0.316^{**} (0.123)	-0.273^{**} (0.119)
Trust			0.603^{***} (0.054)	0.682^{***} (0.052)
Commitment			0.343^{***} (0.05)	0.368^{***} (0.049)
WL 1 (2h vs 5h)	-0.063 (0.268)	0.437^{*} (0.246)		
WL 2 (2h vs 8h)	0.223 (0.276)	0.746^{***} (0.253)		
Late x WL 1	-0.229 (0.385)	-0.473 (0.353)		
Late x WL 2	-0.496 (0.395)	-0.722^{**} (0.363)		
Age	-0.009 (0.006)	0.0100^{*} (0.006)	-0.011^{**} (0.005)	-0.003 (0.004)
Gender	-0.081 (0.169)	-0.017 (0.155)	-0.035 (0.124)	-0.021 (0.120)
Education	0.205 (0.186)	0.152 (0.17)	-0.112 (0.136)	0.033 (0.132)
Income	-0.034 (0.159)	0.053 (0.146)	-0.287^{**} (0.117)	-0.216^{*} (0.113)
Channel	0.230 (0.162)	0.444^{***} (0.148)	-0.067 (0.119)	-0.010 (0.115)
Time orientation	0.010 (0.126)	-0.057 (0.116)	-0.038 (0.091)	-0.114 (0.088)
F-value (*df*)	2.488^{***} (11,303)	5.119^{***} (11,303)	51.123^{***} (9,305)	65.791^{***} (9,305)
R^2	0.288	0.396	0.775	0.812
IMM indirect effect via Trust		IMM WL 1	-0.078 (0.133) CI [-0.356,0.173]	-0.084 (0.141) CI [-0.369,0.182]
		IMM WL 2	-0.170 (0.148) CI [-0.494,0.090]	-0.183 (0.156) CI [-0.508,0.108]
Indirect effect via Trust		IE WL 2 hours	-0.167 (0.100) CI [-0.376,0.020]	-0.180 (0.107) CI [-0.405,0.020]
		IE WL 5 hours	-0.246 (0.103) CI [-0.466,-0.072]	-0.264 (0.107) CI [-0.495,-0.076]
		IE WL 8 hours	-0.337 (0.128) CI [-0.616,-0.121]	-0.362 (0.135) CI [-0.658,-0.126]
IMM indirect effect via Commitment		IMM WL 1	-0.285 (0.219) CI [-0.741,0.121]	-0.323 (0.241) CI [-0.808,0.147]
		IMM WL 2	-0.435 (0.226) CI [-0.897,-0.011]	-0.492 (0.253) CI [-1.004,-0.001]
Indirect effect via Commitment		IE WL 2 hours	-0.279 (0.156) CI [-0.584,0.020]	-0.316 (0.179) CI [-0.678,0.027]
		IE WL 5 hours	-0.565 (0.166) CI [-0.910,-0.263]	-0.639 (0.174) CI [-0.995,-0.304]
		IE WL 8 hours	-0.714 (0.171) CI [-1.074,-0.405]	-0.808 (0.189) CI [-1.179,-0.449]

Note: First stage common to both SAT and REP presented in models 1 and 2. * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$. Standard error in parenthesis.
Abbreviations: IMM = Index of Moderated Mediation; IE = Indirect effect; CI = Confidence Interval.

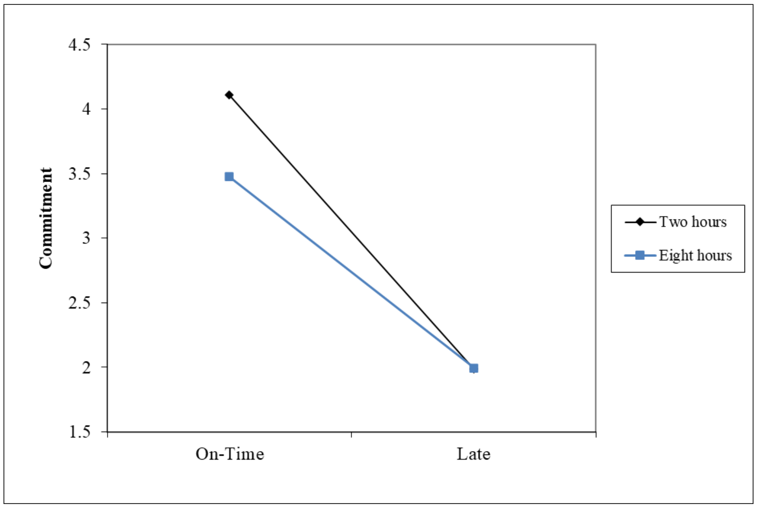


Figure 28 – Interaction effect of late and delivery window length on commitment

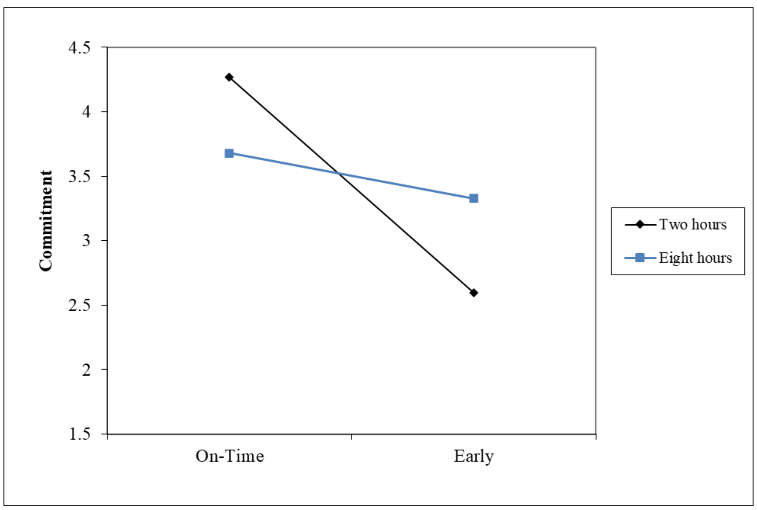


Figure 29 – Interaction effect of late and delivery window length on commitment

Robustness checks

We performed a series of sensitivity analysis to ensure the robustness of our estimations (Appendix F). First, we performed the main analysis without control variables to exclude failed randomization (Lonati et al., 2018). Results confirm the estimations obtained in the main analysis. Hence, the experimental design does not suffer from failed randomization. Second, we ensured that social approval and social desirability did significantly affect the estimations of our model and potentially create demand effect (Lonati et al., 2018). We used multiple strategies and common practices to

address social desirability bias. First, we ensured participants anonymity to exclude any bias related to participants answering the survey to be favorably viewed (Eckerd et al., 2021). Second, we ensured participants that there were no right or wrong answers (Fisher, 1993). Finally, we collected an additional 4-items 7-point (1 = strongly disagree; 7 = strongly agree) scale ($\alpha = 0.81$) adapted from (Fisher, 1993), to assess the extent to which the participant (1) I sometimes feel resentful when I don't get my way, (2) There have been times when I felt like rebelling against people in authority even though I knew they were right, (3) There have been times when I was quite jealous of the good fortune of others, and (4) I am sometimes irritated by people who ask favors of me. We added this scale as additional control to the main analyses and found consistent estimations. Hence, participants' social desirability is not a concern in this study. Third, we performed the analysis including the 77 observations deleted due to failed manipulation check, to avoid any issue related to sample bias. Results reported in Tables 51-52-53-54 provide similar estimations to the main analysis, hence excluding that the 77 removed observations could affect the models.

Study 2 – Discussion of results

The results of Study 2 confirms and expands the results of Study 1. First, Study 2 confirms the negative effect of late and early deliveries on customer outcomes. This suggests that the negative direct effect of early and late is confirmed in terms of both generalizability using the archival dataset, and internal validity using the experiment. Second, this result support the claim from Study 1 that a mediation effect occurs when customers evaluate and compare the service provision with prior expectation. Specifically, a negative disconfirmation decreases the relationship value for the customer, as it erodes trust and commitment. This is especially important to reconcile the contrasting findings relative to repurchase intentions. We claim that customers experiencing an

early delivery will decrease their trust and commitment to the relationship. Hence, while they may still engage in patronage behaviors, their spending will decrease.

Second, Study 2 confirms the negative effect of delivery window on customer outcomes. However, it also shows the difference between an appointment delivery (Two hours) and longer delivery window (Eight hours) for which customers perceive greater uncertainty. As a consequence, customers seem preferring an early delivery instead of facing the uncertainty of a eight-hour delivery window, with better results in terms of commitment, and indirectly to customer outcomes. Overall, when trust is breached, customers do not differentiate between delivery windows as prior expectations will not matter in terms of trusting the service provider.

Third, the post-hoc analysis on reviews informs on customers complaining behaviors (Umashankar et al., 2017). While rating lower level of relationship commitment, customers exposed to early and late conditions would put extra effort to inform the retailer relative to the poor performance, inform other customers, and eventually to correct the next delivery service. Traditionally, literature has linked service failures to negative word-of-mouth (Griffis, Rao, Goldsby, & Niranjan, 2012; Koufteros et al., 2014; Karamana, 2021; Jin et al., 2022). Recent literature suggest that complaining is not necessarily evil for relationship commitment (Umashankar et al., 2017). We show that customers experiencing a service failure would commit less to the relationship, yet offer their effort to inform the retailer.

Conclusions, limitations, and directions for future research

Theoretical Contributions

This study presents three overarching theoretical contributions. First, the importance of customers' evolving expectations on last mile delivery. Past literature emphasized the crucial role of delivery speed (Collier & Bienstock, 2006a; Rao, Goldsby, et al., 2011; Daugherty et al., 2019; M. Fisher

et al., 2019; Akturk et al., 2022). In contrast, recent developments suggest that, as customers become more familiar with home deliveries, they are willing to sacrify service over to gain in service precision (Amorim & DeHoratius, 2021). Our results expand the literature as it offers one of the first studies to find evidence on this growing trend.

Second, we inform prior literature on the trade-off between setting expectations and ensuring the performance. EDT would suggest that managing expectations is of key importance to ensure a positive (or less detrimental) outcome. CTT, instead, would suggest that such expectations impact the relationship but through customers committing to it. Hence, we extend EDT by including the mediation role of commitment onto the relationship. Specifically, not meeting increasing expectations (late delivery combined to a short window or early delivery combined to a longer window) does not inherently produce lower outcomes for the customer-retailer relationship.

Consequently, this study also identifies contribute to the ongoing discussion on customer segmentation relative to the delivery service (Nguyen et al., 2019). With a longer delivery window, late and early deliveries result in a seemingly or better outcome. Promising a shorter window, instead, requires extra effort to deliver within the window. Following operations management literature on travel time predictors (W. Shang & Liu, 2011; Lim et al., 2020, 2023; Stroh et al., 2022), we contribute to this research stream by adding the importance of segmenting based on the promised delivery window. Finally, we also offer contributions in terms of delivery service design. For longer delivery windows, an efficient delivery service may be preferrable, even if it affects customer outcomes, for example in terms of delivery density (Boyer & Hult, 2006; Wang, Rabinovich, & Guda, 2022). Holding that a service failure is detrimental for the customer-retailer

relationship, we inform prior literature that this effect is contingent upon the managing the expectations as well as the relationship with the customer.

Managerial Contributions

This study also offers several managerial contributions. First, it informs managers on the detrimental effects of lateness and earliness, especially in terms of repurchase intentions. In the crowdshipping context, customers expecting punctual deliveries would not repurchase in the same channel after experiencing 40 minutes of lateness. Conversely, earliness would result in a customer return, though also provokes trust erosion and a decrease in commitment. We encourage retailers to design their delivery service to avoid both delivery outcomes, but privilege earliness over lateness when resolving a trade-off on efficiency. For example, assigning multiple orders to the same delivery task will increase density and create efficiency, but it may result in either an early or late order. When retailers face this choice, we encourage them to choose earliness over lateness.

The second overarching managerial contribution results from the recommended customer segmentations based on the promised delivery window. A longer delivery window will give extra slack to designing the delivery service. Retailers may limit the shorter time slots to ensure an on-time delivery on the fewer offered. Alternatively, retailers could offer an alternative channel to shorter delivery slots. Research could further investigate the interplay between service scarcity in terms of fewer shorter delivery slots, and product scarcity in terms of postponing the delivery service to ensure an on-time in-full delivery. This has become even more important in unattended deliveries, where customers positively evaluate greater flexibility given by the longer delivery window (Olsson et al., 2023).

Finally, we remark on the importance, for online retailers, to recognize the evolving relationship with the customer, who provides feedback and commit extra effort to improve the

delivery service. As an integral part of customer shopping experience, last-mile delivery replaces the in-store experience (Daugherty et al., 2019). Online retailers and delivery platforms may accurately learn from customers feedback, as it constitutes the only chance for customers to eventually complain about their experience (Akturk et al., 2022).

Limitations and directions for future research

As with every empirical study, this work presents limitations. First, both studies were focused on the growing market of grocery deliveries performed by crowdshipping platforms. While industry reports show encouraging trends, with an expected growth up to $600 billion for the grocery delivery industry and a growing percentage of customers shifting from click&collect to home deliveries (Chandra et al., 2022), we encourage researchers and practitioners to carefully evaluate the results of this studies and consider the limitations of the dataset and the experimental scenario. For example, a common limitation of these analysis is the impossibility to track customers' purchasing behaviors across channels. For example, a customer may have experienced a late home delivery and never request it again, though being loyal to the retailer. We share this limitation with other studies (Lim et al., 2020), and encourage future research to further investigate cross-channel customers' behaviors (Peinkofer et al., 2022).

Second, despite our efforts to carefully review the last-mile delivery and crowdshipping literature, we focus the moderation effects on only two delivery characteristics. We encourage future research to further explore alternative factors influencing customers prior expectations. Our effort to control for multiple explaining variables can be extended by future research to investigate the moderation effect of, for example, product value and product type (Thirumalai & Sinha, 2005; Nguyen et al., 2019), delivery areas (Merkert et al., 2022), and customers attribution of liability for the delivery failure to which entity of the service chain.

Finally, we investigated these research questions under the lenses of EDT and CTT, two established theories in retail supply chain research (Umashankar et al., 2017; Akturk et al., 2022). Future research could adopt alternative theoretical background to investigate how delivery performance affect customers' emotional reactions though cognitive appraisal theory (Watson & Spence, 2007; Ta et al., 2023), and their satisfaction with the service design, especially in terms of service recovery after experiencing a delivery failure (Dixon & Verma, 2013; Dixon et al., 2017)

Reference list

Abbey, J. D., & Meloy, M. G. (2017). Attention by design: Using attention checks to detect inattentive respondents and improve data quality. *Journal of Operations Management*, *53–56*(1), 63–70. https://doi.org/http://dx.doi.org/10.1016/j.jom.2017.06.001

Abdulla, H., Abbey, J. D., & Ketzenberg, M. (2022). How consumers value retailer's return policy leniency levers: An empirical investigation. *Production and Operations Management*, *15*(32), 4–6. https://doi.org/https://doi.org/10.1111/poms.13640

Agatz, N., Fan, Y., & Stam, D. (2021). The Impact of Green Labels on Time Slot Choice and Operational Sustainability. *Production and Operations Management*, *30*(7), 2285–2303. https://doi.org/10.1111/poms.13368

Akturk, M. S., & Ketzenberg, M. (2022). Exploring the Competitive Dimension of Omnichannel Retailing. *Management Science*, *68*(4), 2732–2750. https://doi.org/10.1287/mnsc.2021.4008

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

Alexander, D., Boone, C., & Lynn, M. (2021). The effects of tip recommendations on customer tipping, satisfaction, repatronage, and spending. *Management Science*, *67*(1), 146–165. https://doi.org/10.1287/mnsc.2019.3541

Ambra, T., Caris, A., & Macharis, C. (2021). Do You See What I See? A Simulation Analysis of Order Bundling within a Transparent User Network in Geographic Space. *Journal of Business Logistics*, *42*(1), 167–190. https://doi.org/10.1111/jbl.12237

Amling, A., & Daugherty, P. J. (2018). Logistics and distribution innovation in China. *International Journal of Physical Distribution and Logistics Management*, *50*(3), 323–332. https://doi.org/10.1108/IJPDLM-07-2018-0273

Amorim, P., & DeHoratius, N. (2021). Online Shoppers Don't Always Care About Faster Delivery. *MIT Sloan Management Review*, *63*(1), 12–14.

Amorim, P., DeHoratius, N., Eng-Larsson, F., & Martins, S. (2020). Customer Preferences for Delivery Service Attributes in Attended Home Delivery. *SSRN Electronic Journal*, *20*. https://doi.org/10.2139/ssrn.3592597

Anderson, E. W. (1998). Customer Satisfaction and Word of Mouth. *Journal of Service Research*, *1*(1), 5–17. https://doi.org/10.1177/109467059800100102

Anderson, E. W., Fornell, C., & Lehmann, D. R. (1994). Customer Satisfaction, Market Share, and Profitability: Findings from Sweden. *Journal of Marketing*, *58*(3), 53. https://doi.org/10.2307/1252310

Anderson, E. W., & Sullivan, M. W. (1993). The Antecedents and Consequences of Customer Satisfaction for Firms. *Marketing Science*, *12*(2), 125–143.

Andritsos, D. A., & Tang, C. S. (2014). Linking process quality and resource usage: An empirical analysis. *Production and Operations Management*, *23*(12), 2163–2177. https://doi.org/10.1111/poms.12249

Arellano, M. C., Meuer, J., & Netland, T. H. (2021). Commitment follows beliefs: A configurational perspective on operations managers' commitment to practice adoption. *Journal of Operations Management*, *67*(4), 450–475. https://doi.org/10.1002/joom.1130

Awaysheh, A., Frohlich, M. T., Flynn, B. B., & Flynn, P. J. (2021). To err is human: Exploratory multilevel analysis of supply chain delivery delays. *Journal of Operations Management*, *67*(7), 882–916. https://doi.org/https://doi.org/10.1002/joom.1154

Azadegan, A., Patel, P. C., & Parida, V. (2013). Operational slack and venture survival. *Production and Operations Management*, *22*(1), 1–18. https://doi.org/10.1111/j.1937-5956.2012.01361.x

Bachrach, D. G., & Bendoly, E. (2011). Rigor in behavioral experiments: A basic primer for supply chain management researchers. *Journal of Supply Chain Management*, *47*(3), 5–8. https://doi.org/https://doi.org/10.1111/j.1745-493X.2011.03230.x

Bagozzi, R. P., & Yi, Y. (1988). On the evaluation of structural equation models. *Journal of the Academy of Marketing Science*, *16*(1), 74–94. https://doi.org/10.1007/BF02723327

Balasubramanian, S. (1998). Mail versus mall: A strategic analysis of competition between direct marketers and conventional retailers. *Marketing Science*, *17*(3), 181–195. https://doi.org/10.1287/mksc.17.3.181

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

Barrios, J. M., Hochberg, Y. V., & Yi, H. (2022). The cost of convenience: Ridehailing and traffic fatalities. *Journal of Operations Management*, *September 2021*, 1–33. https://doi.org/10.1002/joom.1221

Belenzon, S., Chatterji, A. K., & Daley, B. (2020). Choosing between growth and glory. *Management Science*, *66*(5), 2050–2074. https://doi.org/10.1287/mnsc.2019.3296

Belo, R., Ferreira, P., & Telang, R. (2014). Broadband in school: Impact on student performance. *Management Science*, *60*(2), 265–282. https://doi.org/10.1287/mnsc.2013.1770

Berry, L. L., Seiders, K., & Grewal, D. (2002). Understanding Service Convenience. *Journal of Marketing*, *66*(3), 1–17. https://doi.org/https://doi.org/10.1509/jmkg.66.3.1.18505

Bhan, H. S. (Zack), & Anderson, E. T. (2023). Multiyear Impact of Backorder Delays: A Quasi-Experimental Approach. *Marketing Science*, *42*(2), 314–335. https://doi.org/10.1287/mksc.2022.1384

Bhatia, S. (2019). Predicting risk perception: New insights from data science. *Management Science*, *65*(8), 3800–3823. https://doi.org/10.1287/mnsc.2018.3121

Bhattacharjee, S., Gopal, R. D., Lertwachara, K., Marsden, J. R., & Telang, R. (2007). The effect of digital sharing technologies on music markets: A survival analysis of albums on ranking charts. *Management Science*, *53*(9), 1359–1374. https://doi.org/10.1287/mnsc.1070.0699

Bhattacherjee, A. (2001). Understanding Information Systems Continuance: An Expectation-Confirmation Model. *MIS Quarterly*, *25*(3), 351. https://doi.org/10.2307/3250921

Bitner, M. J. (1990). Evaluating Service Encounters: The Effects of Physical Surroundings and Employee Responses. *Journal of Marketing*, *54*(2), 69–82. https://doi.org/10.2307/1251871

Bitner, M. J. (1995). Building Service Relationships: It's all about Promises. *Journal of the Academy of Marketing Science*, *23*(4), 246–251. https://doi.org/10.1177/009207039502300403

Bitran, G. R., Ferrer, J. C., & E Oliveira, P. R. (2008). Managing customer experiences: Perspectives on the temporal aspects of service encounters. *Manufacturing and Service Operations Management*, *10*(1), 61–83. https://doi.org/10.1287/msom.1060.0147

Bode, C., Lindemann, E., & Wagner, S. M. (2011). Driving trucks and driving sales? the impact of delivery personnel on customer purchase behavior. *Journal of Business Logistics*, *32*(1), 99–114. https://doi.org/https://doi.org/10.1111/j.2158-1592.2011.01009.x

Bolton, R. N. (1998). A dynamic model of the duration of the customer's relationship with a continuous service provider: The role of satisfaction. *Marketing Science*, *17*(1), 45–65. https://doi.org/10.1287/mksc.17.1.45

Boyer, K. K., & Hult, G. T. M. (2006). Customer behavioral intentions for online purchases: An examination of fulfillment method and customer experience level. *Journal of Operations Management*, *24*(2), 124–147. https://doi.org/https://doi.org/10.1016/j.jom.2005.04.002

Boyer, K. K., Prud'homme, A. M., & Chung, W. (2009). The Last Mile Challenge: Evaluating the Effects Of Customer Density and Delivery Window Patterns. *Journal of Business Logistics*, *30*(1), 185–201. https://doi.org/10.1002/j.2158-1592.2009.tb00104.x

Bradburn, M. J., Clark, T. G., Love, S. B., & Altman, D. G. (2003). Survival Analysis Part II: Multivariate data analysis- An introduction to concepts and methods. *British Journal of Cancer*, *89*(3), 431–436. https://doi.org/10.1038/sj.bjc.6601119

Brown, S. A., Venkatesh, V., Kuruzovich, J., & Massey, A. P. (2008). Expectation confirmation:

An examination of three competing models. *Organizational Behavior and Human Decision Processes*, *105*(1), 52–66. https://doi.org/10.1016/j.obhdp.2006.09.008

Buell, R. W., Campbell, D., & Frei, F. X. (2010). Are self-service customers satisfied or stuck? *Production and Operations Management*, *19*(6), 679–697. https://doi.org/10.1111/j.1937-5956.2010.01151.x

Campbell, A. M., & Savelsbergh, M. W. P. (2005). Decision support for consumer direct grocery initiatives. *Transportation Science*, *39*(3), 313–327. https://doi.org/10.1287/trsc.1040.0105

Caro, F., Kök, A. G., & Martínez-de-Albéniz, V. (2020). The Future of Retail Operations. *Manufacturing & Service Operations Management*, *22*(1), 47–58. https://doi.org/10.1287/msom.2019.0824

Castillo, V. E., Mollenkopf, D. A., Bell, J. E., & Esper, T. L. (2022). Designing technology for on-demand delivery: The effect of customer tipping on crowdsourced driver behavior and last mile performance. *Journal of Operations Management*, *68*(5), 424–453. https://doi.org/10.1002/joom.1187

Chandra, V., Gill, P., Kohli, S., Mathur, V., Venkataraman, K., & Yoshimura, J. (2022). *The next horizon for grocery e-commerce: Beyond the pandemic bump*. McKinsey. https://www.mckinsey.com/industries/retail/our-insights/the-next-horizon-for-grocery-ecommerce-beyond-the-pandemic-bump

Chang, Y. Y., Dasgupta, S., & Hilary, G. (2010). CEO ability, pay, and firm performance. *Management Science*, *56*(10), 1633–1652. https://doi.org/10.1287/mnsc.1100.1205

Chen, P.-Y., Hong, Y., & Liu, Y. (2018). The Value of Multidimensional Rating Systems: Evidence from a Natural Experiment and Randomized Experiments. *Management Science*, *64*(10), 4629–4647. https://doi.org/10.1287/mnsc.2017.2852

Chen, Y., Fay, S., & Wang, Q. (2011). The Role of Marketing in Social Media: How Online Consumer Reviews Evolve. *Journal of Interactive Marketing*, *25*(2), 85–94. https://doi.org/10.1016/j.intmar.2011.01.003

Choudhary, V., Shunko, M., Netessine, S., & Koo, S. (2021). Nudging Drivers to Safety: Evidence from a Field Experiment. *Management Science*. https://doi.org/10.1287/mnsc.2021.4063

Clark, T. G., Bradburn, M. J., Love, S. B., & Altman, D. G. (2003). Survival Analysis Part I: Basic concepts and first analyses. *British Journal of Cancer*, *89*(2), 232–238. https://doi.org/10.1038/sj.bjc.6601118

Collier, J. E., & Bienstock, C. C. (2006a). How do customers judge quality in an E-tailer? *MIT Sloan Management Review*, *48*(1), 35–40.

Collier, J. E., & Bienstock, C. C. (2006b). Measuring service quality in E-retailing. *Journal of Service Research*, *8*(3), 260–275. https://doi.org/10.1177/1094670505278867

Confente, I., Russo, I., Peinkofer, S., & Frankel, R. (2021). The challenge of remanufactured products: the role of returns policy and channel structure to reduce consumers' perceived risk. *International Journal of Physical Distribution & Logistics Management*, *51*(4), 350–380. https://doi.org/10.1108/IJPDLM-03-2020-0089

Corbett, C. J., Montes-Sancho, M. J., & Kirsch, D. A. (2005). The financial impact of ISO 9000 certification in the United States: An empirical analysis. *Management Science*, *51*(7), 1046–1059. https://doi.org/10.1287/mnsc.1040.0358

Cox, D. R. (1972). Regression Models and Life-Tables. *Journal of the Royal Statistical Society: Series B (Methodological)*, *34*(2), 187–202. https://doi.org/10.1111/j.2517-6161.1972.tb00899.x

Craighead, C. W., Ketchen, D. J., Dunn, K. S., & Hult, G. T. M. (2011). Addressing common method variance: Guidelines for survey research on information technology, operations, and supply chain management. *IEEE Transactions on Engineering Management*, *58*(3), 578–588. https://doi.org/https://doi.org/10.1109/TEM.2011.2136437

Cui, R., Li, M., & Li, Q. (2020). Value of high-quality logistics: Evidence from a clash between SF express and alibaba. *Management Science*, *66*(9), 3879–3902. https://doi.org/10.1287/mnsc.2019.3411

Cui, R., Sun, T., Lu, Z., & Golden, J. (2020). Sooner or Later? Promising Delivery Speed in Online Retail. *SSRN Electronic Journal*, *April*. https://doi.org/10.2139/ssrn.3563404

Dabholkar, P. A., Thorpe, D. I., & Rentz, J. O. (1996). A measure of service quality for retail stores: Scale development and validation. *Journal of the Academy of Marketing Science*, *24*(1), 3–16. https://doi.org/10.1007/BF02893933

Daugherty, P. J., Bolumole, Y., & Grawe, S. J. (2019). The new age of customer impatience: An agenda for reawakening logistics customer service research. *International Journal of Physical Distribution & Logistics Management*, *49*(1), 4–32. https://doi.org/https://doi.org/10.1108/ IJPDLM-03-2018-0143

Davis-Sramek, B., Droge, C., Mentzer, J. T., & Myers, M. B. (2009). Creating commitment and loyalty behavior among retailers: What are the roles of service quality and satisfaction? *Journal of the Academy of Marketing Science*, *37*(4), 440–454. https://doi.org/10.1007/s11747-009-0148-y

Davis-Sramek, B., Germain, R., & Stank, T. P. (2010). the Impact of Order Fulfillment Service on Retailer Merchandising Decisions in the Consumer Durables Industry. *Journal of Business Logistics*, *31*(2), 215–230. https://doi.org/10.1002/j.2158-1592.2010.tb00149.x

Davis-Sramek, B., Mentzer, J. T., & Stank, T. P. (2008). Creating consumer durable retailer customer loyalty through order fulfillment service operations. *Journal of Operations Management*, *26*(6), 781–797. https://doi.org/https://doi.org/10.1016/j.jom.2007.07.001

Dayarian, I., & Pazour, J. (2022). Crowdsourced order-fulfillment policies using in-store customers. *Production and Operations Management*, 1–32. https://doi.org/10.1111/poms.13805

Dayarian, I., & Savelsbergh, M. (2020). Crowdshipping and Same-day Delivery: Employing In-store Customers to Deliver Online Orders. *Production and Operations Management*, *29*(9), 2153–2174. https://doi.org/10.1111/poms.13219

De Vries, J., DeKoster, R., & Stam, D. (2016). Aligning Order Picking Methods, Incentive Systems, and Regulatory Focus to Increase Performance. *Production and Operations*

Management, *25*(8), 1363–1376. https://doi.org/10.1111/poms.12547

De Vries, J., Roy, D., & De Koster, R. (2018). Worth the wait? How restaurant waiting time influences customer behavior and revenue. *Journal of Operations Management*, *63*(May), 59–78. https://doi.org/10.1016/j.jom.2018.05.001

Deng, Q., Fang, X., & Lim, Y. F. (2021). Urban Consolidation Center or Peer-to-Peer Platform? The Solution to Urban Last-Mile Delivery. *Production and Operations Management*, *30*(4), 997–1013. https://doi.org/10.1111/poms.13289

Dhanorkar, S., & Siemsen, E. (2021). How Nudges Lead to Improved Energy Efficiency in Manufacturing: Evidence from Archival Data and a Field Study. *Production and Operations Management*, *30*(10), 3735–3757. https://doi.org/10.1111/poms.13479

Dixon, M., & Thompson, G. M. (2019). The Impact of Timing and Bundling Flexibility on Affect-Based Service Package Design. *Decision Sciences*, *50*(5), 948–984. https://doi.org/10.1111/deci.12361

Dixon, M., & Verma, R. (2013). Sequence effects in service bundles: Implications for service design and scheduling. *Journal of Operations Management*, *31*(3), 138–152. https://doi.org/10.1016/j.jom.2012.12.002

Dixon, M., Victorino, L., Kwortnik, R. J., & Verma, R. (2017). Surprise, Anticipation, and Sequence Effects in the Design of Experiential Services. *Production and Operations Management*, *26*(5), 945–960. https://doi.org/10.1111/poms.12675

Dobrzykowski, D. D., Mcfadden, K. L., & Vonderembse, M. A. (2020). Examining pathways to safety and financial performance in hospitals : A study of lean in professional service operations. *Journal of Operations Management*, *42–43*(2016), 39–51. https://doi.org/10.1016/j.jom.2016.03.001

Donohue, K., Özer, Ö., & Zheng, Y. (2020). Behavioral operations: Past, present, and future. *Manufacturing and Service Operations Management*, *22*(1), 191–202. https://doi.org/10.1287/msom.2019.0828

Doordash.com. (2023). *Lateness-Based Violations Explained*. Doordash.Com. https://help.doordash.com/dashers/s/article/Lateness-Based-Deactivations-Explained?language=en_US#:~:text=A delivery is considered late,%2C wait time at merchant).

Eckerd, S., DuHadway, S., Bendoly, E., Carter, C., & Kaufmann, L. (2021). On making experimental design choices: Discussions on the use and challenges ofdemand effects, incentives, deception, samples, and vignettes. *Journal of Operations Management*, *67*(2), 261–275. https://doi.org/https://doi.org/10.1002/joom.1128

Esper, T. L. (2021). Supply Chain Management Amid the Coronavirus Pandemic. *Journal of Public Policy & Marketing*, *40*(1), 101–102. https://doi.org/https://doi.org/10.1177/0743915620932150

Esper, T. L., Jensen, T. D., Turnipseed, F. L., & Burton, S. (2003). The Last Mile: an Examination of Effects of Online Retail Delivery Strategies on Consumers. *Journal of Business Logistics*, *24*(2), 177–203. https://doi.org/https://doi.org/10.1002/j.2158-1592.2003.tb00051.x

Farber, H. S. (2015). Why you can't find a taxi in the rain and other labor supply lessons from cab drivers. *Quarterly Journal of Economics*, *130*(4), 1975–2026. https://doi.org/10.1093/qje/qjv026

Fatehi, S., & Wagner, M. R. (2022). Crowdsourcing Last-Mile Deliveries. *Manufacturing & Service Operations Management*, *24*(2), 791–809. https://doi.org/10.1287/msom.2021.0973

Fisher, M., Gallino, S., & Xu, J. J. (2015). The Value of Rapid Delivery in Online Retailing. *SSRN Electronic Journal*, *56*(5), 732–748. https://doi.org/10.2139/ssrn.2573069

Fisher, M., Gallino, S., & Xu, J. J. (2019). The Value of Rapid Delivery in Omnichannel Retailing. *Journal of Marketing Research*, *56*(5), 732–748. https://doi.org/10.1177/0022243719849940

Fisher, R. (1993). Social Desirability Bias and the Validity of Indirect Questioning. *Journal of Consumer Research*, *20*(2), 303. https://doi.org/10.1086/209351

Flynn, R. (2012). Survival analysis. *Journal of Clinical Nursing*, *21*(19–20), 2789–2797. https://doi.org/10.1111/j.1365-2702.2011.04023.x

Fornell, C., & Larcker, D. F. (1981). Evaluating Structural Equation Models with Unobservable Variables and Measurement Error. *Journal of Marketing Research*, *18*(1), 39–50. https://doi.org/https://doi.org/10.2307/3151312

Franke, H., Eckerd, S., & Foerstl, K. (2022). Rising to the Top: Motivational Forces Influencing Status Conflict in Sourcing Teams. *Production and Operations Management*, *31*(3), 963–983. https://doi.org/10.1111/poms.13590

Gambeta, E., Koka, B. R., & Hoskisson, R. E. (2019). Being too good for your own good: A stakeholder perspective on the differential effect of firm-employee relationships on innovation search. *Strategic Management Journal*, *40*(1), 108–126. https://doi.org/10.1002/smj.2967

Garbarino, E., & Johnson, M. S. (1999). The Different Roles of Satisfaction, Trust, and Commitment in Customer Relationships. *Journal of Marketing*, *63*(2), 70–87. https://doi.org/10.1177/002224299906300205

Gawor, T., & Hoberg, K. (2019). Customers' valuation of time and convenience in e-fulfillment. *International Journal of Physical Distribution and Logistics Management*, *49*(1), 75–98. https://doi.org/10.1108/IJPDLM-09-2017-0275

Golicic, S. L., & Davis, D. F. (2012). Implementing mixed methods research in supply chain management. *International Journal of Physical Distribution & Logistics Management*, *42*(8/9), 726–741. https://doi.org/10.1108/09600031211269721

Gordon, G., Greenwood, B. N., Agarwal, R., & Jeffrey, S. (2015). Vocal Minority and Silent Majority: How Do Online Ratings Reflect Population Perceptions of Quality. *MIS Quarterly*, *39*(3), 565–590.

Griffis, S. E., Rao, S., Goldsby, T. J., & Niranjan, T. T. (2012). The customer consequences of returns in online retailing: An empirical analysis. *Journal of Operations Management*, *30*(4), 282–294. https://doi.org/10.1016/j.jom.2012.02.002

Griffis, S. E., Rao, S., Goldsby, T. J., Voorhees, C. M., & Iyengar, D. (2012). Linking order fulfillment performance to referrals in online retailing: An empirical analysis. *Journal of Business Logistics*, *33*(4), 279–294. https://doi.org/https://doi.org/10.1111/jbl.12002

Grout, J. R. (1998). Influencing a Supplier Using Delivery Windows: Its Effect on the Variance of Flow Time and On-Time Delivery. *Decision Sciences*, *29*(3), 747–764. https://doi.org/10.1111/j.1540-5915.1998.tb01362.x

Guda, H., & Subramaniana, U. (2019). Your uber is arriving: Managing on-demand workers through surge pricing, forecast communication, and worker incentives. *Management Science*, *65*(5), 1995–2014. https://doi.org/10.1287/mnsc.2018.3050

Gümüş, M., Li, S., Oh, W., & Ray, S. (2013). Shipping fees or shipping free? a tale of two price partitioning strategies in online retailing. *Production and Operations Management*, *22*(4), 758–776. https://doi.org/10.1111/j.1937-5956.2012.01391.x

Guo, F., & Liu, Y. (2023). EXPRESS: The Effectiveness of Membership-based Free Shipping: An Empirical Investigation on Consumers' Purchase Behaviors and Revenue Contribution. *Journal of Marketing*. https://doi.org/10.1177/00222429231158116

Han, S., Chen, S., Yang, K., Li, H., Yang, F., & Luo, Z. (2022). Free shipping policy for imported cross-border e-commerce platforms. *Annals of Operations Research*. https://doi.org/10.1007/s10479-022-04774-5

Hayes, A. F. (2018). *Introduction to Mediation, Moderation, and Conditional Process Analysis: A Regression-Based Approach*. Guilford Press.

Hayes, A. F., & Montoya, A. K. (2017). A Tutorial on Testing, Visualizing, and Probing an Interaction Involving a Multicategorical Variable in Linear Regression Analysis. *Communication Methods and Measures*, *11*(1), 1–30. https://doi.org/10.1080/19312458.2016.1271116

Heckman, J. J. (1979). Sample Selection Bias as a Specification Error. *Econometrica*, *47*(1), 153. https://doi.org/10.2307/1912352

Hendricks, K. B., & Singhal, V. R. (2014). The effect of demand-supply mismatches on firm risk. *Production and Operations Management*, *23*(12), 2137–2151. https://doi.org/10.1111/poms.12084

Hennig-Thurau, T., Groth, M., Paul, M., & Gremler, D. D. (2006). Are All Smiles Created Equal? How Emotional Contagion and Emotional Labor Affect Service Relationships. *Journal of Marketing*, *70*(3), 58–73. https://doi.org/https://doi.org/10.1509/jmkg.70.3.058

Hess, R. L., Ganesan, S., & Klein, N. M. (2003). Service failure and recovery: The impact of relationship factors on customer satisfaction. *Journal of the Academy of Marketing Science*, *31*(2), 127–145. https://doi.org/https://doi.org/10.1177/0092070302250898

Ho, T. H., Lim, N., Reza, S., & Xia, X. (2017). OM forum - Causal inference models in operations management. *Manufacturing and Service Operations Management*, *19*(4), 509–525. https://doi.org/10.1287/msom.2017.0659

Ho, T. H., & Zheng, Y. S. (2004). Setting customer expectation in service delivery: An integrated

marketing-operations perspective. *Management Science*, *50*(4), 479–488. https://doi.org/10.1287/mnsc.1040.0170

Hogarth, R. M., & Einhorn, H. J. (1992). Order effects in belief updating: The belief-adjustment model. *Cognitive Psychology*, *24*(1), 1–55. https://doi.org/10.1016/0010-0285(92)90002-J

Hosmer, D., & Lemeshow, S. (1999). *Applied survival analysis: regression modeling of time to event data*. JohnWiley and sons.

Hu, L. T., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling*, *6*(1), 1–55. https://doi.org/https://doi.org/10.1080/10705519909540118

Hübner, A., Kuhn, H., & Wollenburg, J. (2016). Last mile fulfilment and distribution in omni-channel grocery retailing: A strategic planning framework. *International Journal of Retail and Distribution Management*, *44*(3), 228–247. https://doi.org/10.1108/IJRDM-11-2014-0154

Ishfaq, R., Defee, C. C., Gibson, B. J., & Raja, U. (2016). Realignment of the physical distribution process in omni-channel fulfillment. *International Journal of Physical Distribution and Logistics Management*, *46*(6–7), 543–561. https://doi.org/10.1108/IJPDLM-02-2015-0032

Jin, Y. "Henry," Ueltschy Murfield, M. L., & Bock, D. E. (2022). Do as You Say, or I Will: Retail signal congruency in buy-online-pickup-in-store and negative word-of-mouth. *Journal of Business Logistics*, *July*, 1–24. https://doi.org/https://doi.org/10.1111/jbl.12322

Kamalahmadi, M., Yu, Q., & Zhou, Y. P. (2021). Call to Duty: Just-in-Time Scheduling in a Restaurant Chain. *Management Science*, *67*(11), 6751–6781. https://doi.org/10.1287/mnsc.2020.3877

Kammerer, W., Chinnareddy, S., Mahoney, K., & Neff, D. (2020). *The last mile: Enabling last mile through speed and flexibility upstream*.

Kaplan, E. L., & Meier, P. (1958). Nonparametric Estimation from Incomplete Observations. *Journal of the American Statistical Association*, *53*(282), 457–481.

Karamana, H. (2021). Online review solicitations reduce extremity bias in online review distributions and increase their representativeness. *Management Science*, *67*(7), 4420–4445. https://doi.org/10.1287/mnsc.2020.3758

Kaufman, C. F., Lane, P. M., & Lindquist, J. D. (1991). Exploring More than 24 Hours a Day: A Preliminary Investigation of Polychronic Time Use. *Journal of Consumer Research*, *18*(3), 392. https://doi.org/10.1086/209268

Ketchen, D. J., Kaufmann, L., & Carter, C. R. (2021). Configurational approaches to theory development in supply chain management: Leveraging underexplored opportunities. *Journal of Supply Chain Management*, *August*, 1–18. https://doi.org/https://doi.org/10.1111/jscm.12275

Khan, R., Lewis, M., & Singh, V. (2009). Dynamic customer management and the value of one-to-one marketing. *Marketing Science*, *28*(6), 1063–1079. https://doi.org/10.1287/mksc.1090.0497

Kokkodis, M., Lappas, T., & Kane, G. C. (2022). Optional purchase verification in e-commerce platforms: More representative product ratings and higher quality reviews. *Production and Operations Management*, *Forthcoming*. https://doi.org/10.1111/poms.13731

Koufteros, X., Droge, C., Heim, G., Massad, N., & Vickery, S. K. (2014). Encounter Satisfaction in E-tailing: Are the relationships of order fulfillment service quality with its antecedents and consequences moderated by historical satisfaction? *Decision Sciences*, *45*(1), 5–48. https://doi.org/https://doi.org/10.1111/deci.12056

Kwon, I. W. G., & Suh, T. (2004). Factors affecting the level of trust and commitment in supply chain relationships. *Journal of Supply Chain Management*, *40*(1), 4–14. https://doi.org/10.1111/j.1745-493X.2004.tb00165.x

Lee, H. L., & Whang, S. (2001). Winning the Last Mile of E-Commerce. *MIT Sloan Management Review*, *42*(4), 54–62.

Lee, M., Verma, R., & Roth, A. (2015). Understanding customer value in technology-enabled services: A numerical taxonomy based on usage and utility. *Service Science*, *7*(3), 227–248. https://doi.org/10.1287/serv.2015.0105

Leng, M., & Becerril-Arreola, R. (2010). Joint pricing and contingent free-shipping decisions in B2C transactions. *Production and Operations Management*, *19*(4), 390–405. https://doi.org/10.1111/j.1937-5956.2009.01112.x

Lewis, M., Singh, V., & Fay, S. (2006). An empirical study of the impact of nonlinear shipping and handling fees on purchase incidence and expenditure decisions. *Marketing Science*, *25*(1), 51–64. https://doi.org/https://doi.org/10.1287/mksc.1050.0150

Li, M., Choi, T. Y., Rabinovich, E., & Crawford, A. (2013). Self-service operations at retail stores: The role of inter-customer interactions. *Production and Operations Management*, *22*(4), 888–914. https://doi.org/10.1111/j.1937-5956.2012.01321.x

Li, S., Rao, S. S., Ragu-Nathan, T. S., & Ragu-Nathan, B. (2005). Development and validation of a measurement instrument for studying supply chain management practices. *Journal of Operations Management*, *23*(6), 618–641. https://doi.org/https://doi.org/10.1016/j.jom.2005.01.002

Lim, S. F. W. T., Gao, F., & Tan, T. (2020). Slow and Sure, or Fast and Furious? An Empirical Study About Omnichannel Demand Sensitivity to Fulfillment Lead Time. *SSRN Electronic Journal*, 1–63. https://doi.org/10.2139/ssrn.3544824

Lim, S. F. W. T., Jin, X., & Srai, J. S. (2018). Consumer-driven e-commerce: A literature review, design framework, and research agenda on last-mile logistics models. *International Journal of Physical Distribution and Logistics Management*, *48*(3), 308–332. https://doi.org/10.1108/IJPDLM-02-2017-0081

Lim, S. F. W. T., Wang, Q., & Webster, S. (2023). Do it right the first time: Vehicle routing with home delivery attempt predictors. *Production and Operations Management*, *Forthcomin*. https://doi.org/10.1111/poms.13926

Lim, S. F. W. T., & Winkenbach, M. (2019). Configuring the last-mile in businessto- consumer e-retailing. *California Management Review*, *61*(2), 132–154.

https://doi.org/https://doi.org/10.1177/0008125618805094

Liu, S., He, L., & Shen, Z. J. M. (2021). On-time last-mile delivery: Order assignment with travel-time predictors. *Management Science*, *67*(7), 4095–4119. https://doi.org/10.1287/mnsc.2020.3741

Lonati, S., Quiroga, B. F., Zehnder, C., & Antonakis, J. (2018). On doing relevant and rigorous experiments: Review and recommendations. *Journal of Operations Management*, *64*(1), 19–40. https://doi.org/https://doi.org/10.1016/j.jom.2018.10.003

Lu, G., Ding, X. (David), Peng, D. X., & Hao-Chun Chuang, H. (2018). Addressing endogeneity in operations management research: Recent developments, common problems, and directions for future research. *Journal of Operations Management*, *64*(1), 53–64. https://doi.org/10.1016/j.jom.2018.10.001

Luo, X., Tong, S., Lin, Z., & Zhang, C. (2021). The Impact of Platform Protection Insurance on Buyers and Sellers in the Sharing Economy: A Natural Experiment. *Journal of Marketing*, *85*(2), 50–69. https://doi.org/10.1177/0022242920962510

Ma, S. (2017). Fast or free shipping options in online & Omni-channel retail? The mediating role of uncertainty on satisfaction & purchase intentions. *International Journal of Logistics Management*, *28*(4), 1099–1122. https://doi.org/https://doi.org/10.1108/IJLM-05-2016-0130

Mangiaracina, R., Perego, A., Seghezzi, A., & Tumino, A. (2019). Innovative solutions to increase last-mile delivery efficiency in B2C e-commerce: a literature review. *International Journal of Physical Distribution and Logistics Management*, *49*(9), 901–920. https://doi.org/https://doi.org/10.1108/IJPDLM-02-2019-0048

Mao, W., Ming, L., Rong, Y., Tang, C. S., & Zheng, H. (2019). Faster Deliveries and Smarter Order Assignments for an On-Demand Meal Delivery Platform. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.3469015

Mentzer, J., Flint, D., & Kent, J. (1999). Developing a logistics service quality scale. *Journal of Business Logistics*, *20*(1), 9.

Mentzer, J. T., Flint, D. J., & Kent, J. L. (1999). Developing a logistics service quality scale. *Journal of Business Logistics*, *9*(2), 1–47.

Merkert, R., Bliemer, M. C. J., & Fayyaz, M. (2022). Consumer preferences for innovative and traditional last-mile parcel delivery. *International Journal of Physical Distribution & Logistics Management*, *52*(3), 261–284. https://doi.org/10.1108/ijpdlm-01-2021-0013

Miao, W., Deng, Y., Wang, W., Liu, Y., & Tang, C. S. (2022). The effects of surge pricing on driver behavior in the ride-sharing market: Evidence from a quasi-experiment. *Journal of Operations Management, Forthcomin*. https://doi.org/10.1002/joom.1223

Miller, J. W., Davis-Sramek, B., Fugate, B. S., Pagell, M., & Flynn, B. B. (2020). Editorial commentary: Addressing confusion in the diffusion of archival data research. In *Journal of Supply Chain Management*. https://doi.org/10.1111/jscm.12236

Miller, J. W., Stromeyer, W. R., & Schwieterman, M. A. (2013). Extensions of the Johnson-Neyman Technique to Linear Models With Curvilinear Effects: Derivations and Analytical

Tools. *Multivariate Behavioral Research*, *48*(2), 267–300. https://doi.org/10.1080/00273171.2013.763567

Mithas, S., Chen, Y., Lin, Y., & De Oliveira Silveira, A. (2022). On the causality and plausibility of treatment effects in operations management research. *Production and Operations Management*, *31*(12), 4558–4571. https://doi.org/10.1111/poms.13863

Mittal, V., Ross, W. T., & Baldasare, P. M. (1998). The asymmetric impact of negative and positive attribute-level performance on overall satisfaction and repurchase intentions. *Journal of Marketing*, *62*(1), 33–47. https://doi.org/https://doi.org/10.2307/1251801

Mollenkopf, D. A., Peinkofer, S. T., & Chu, Y. (J. (2022). Supply chain transparency: Consumer reactions to incongruent signals. *Journal of Operations Management*, *68*(4), 306–327. https://doi.org/https://doi.org/10.1002/joom.1180

Morgan, R. M., & Hunt, S. D. (1994). The Commitment-Trust Theory of Relationship Marketing. *Journal of Marketing*, *58*(3), 20. https://doi.org/10.2307/1252308

Murfield, M., Boone, C. A., Rutner, P., & Thomas, R. (2017). Investigating logistics service quality in omni-channel retailing. *International Journal of Physical Distribution and Logistics Management*, *47*(4), 263–296. https://doi.org/10.1108/IJPDLM-06-2016-0161

Nguyen, D. H., de Leeuw, S., & Dullaert, W. E. H. (2018). Consumer Behaviour and Order Fulfilment in Online Retailing: A Systematic Review. *International Journal of Management Reviews*, *20*(2), 255–276. https://doi.org/10.1111/ijmr.12129

Nguyen, D. H., de Leeuw, S., Dullaert, W., & Foubert, B. P. J. (2019). What Is the Right Delivery Option for You? Consumer Preferences for Delivery Attributes in Online Retailing. *Journal of Business Logistics*, *40*(4), 299–321. https://doi.org/https://doi.org/10.1111/jbl.12210

Nichols, B. S., Stolze, H., & Kirchoff, J. F. (2019). Spillover effects of supply chain news on consumers' perceptions of product quality: An examination within the triple bottom line. *Journal of Operations Management*, *65*(6), 536–559. https://doi.org/https://doi.org/10.1002/joom.1033

Nunnally, J. C., & Bernstein, I. (1994). *Psychometric Theory* (3rd Editio). McGraw-Hil.

Oliver, R. L. (1977). Effect of expectation and disconfirmation on postexposure product evaluations: An alternative interpretation. *Journal of Applied Psychology*, *62*(4), 480.

Oliver, R. L. (1980). A Cognitive Model of the Antecedents and Consequences of Satisfaction Decisions. *Journal of Marketing Research*, *17*(4), 460–469. https://doi.org/10.2307/3150499

Oliver, R. L. (1993). Cognitive, Affective, and Attribute Bases of the Satisfaction Response. *Journal of Consumer Research*, *20*(3), 418. https://doi.org/https://doi.org/10.1086/209358

Olson, J. C., & Dover, P. A. (1979). Disconfirmation of consumer expectations through product trial. *Journal of Applied Psychology*, *64*(2), 179–189. https://doi.org/10.1037/0021-9010.64.2.179

Olsson, J., Hellström, D., & Vakulenko, Y. (2023). Customer experience dimensions in last-mile delivery: an empirical study on unattended home delivery. *International Journal of Physical*

Distribution & Logistics Management, *53*(2), 184–205. https://doi.org/10.1108/IJPDLM-12-2021-0517

Peinkofer, S. T., Esper, T. L., & Howlett, E. (2016). Hurry! Sale Ends Soon: The Impact of Limited Inventory Availability Disclosure on Consumer Responses to Online Stockouts. *Journal of Business Logistics*, *37*(3), 231–246. https://doi.org/https://doi.org/10.1111/jbl.12136

Peinkofer, S. T., Esper, T. L., Smith, R. J., & Williams, B. D. (2022). Retail "Save the Sale" tactics: Consumer perceptions of in-store logistics service recovery. *Journal of Business Logistics*, *43*(2), 238–264. https://doi.org/https://doi.org/10.1111/jbl.12294

Peinkofer, S. T., & Jin, Y. H. (2022). The impact of order fulfillment information disclosure on consequences of deceptive counterfeits. *Production and Operations Management*, *15*(2), 1–23. https://doi.org/https://doi.org/10.1111/poms.13833

Peinkofer, S. T., Schwieterman, M. A., & Miller, J. W. (2020). Last-mile delivery in the motor-carrier industry: A panel data investigation using discrete time event history analysis. *Transportation Journal*, *59*(3), 129–164. https://doi.org/https://doi.org/10.5325/transportationj.59.2.0129

Peng, D. X., & Lu, G. (2017). Exploring the Impact of Delivery Performance on Customer Transaction Volume and Unit Price: Evidence from an Assembly Manufacturing Supply Chain. *Production and Operations Management*, *26*(5), 880–902. https://doi.org/10.1111/poms.12682

Perdikaki, O., Peng, D. X., & Heim, G. R. (2015). Impact of Customer Traffic and Service Process Outsourcing Levels on e-Retailer Operational Performance. *Production and Operations Management*, *24*(11), 1794–1811. https://doi.org/10.1111/poms.12359

Picard, R. (2022). *help geodist*. http://fmwww.bc.edu/repec/bocode/g/geodist.html

Praet, S., & Martens, D. (2020). Efficient Parcel Delivery by Predicting Customers' Locations. *Decision Sciences*, *51*(5), 1202–1231. https://doi.org/10.1111/deci.12376

Rabinovich, E. (2004). Internet Retailing Intermediation: a Multilevel Analysis of Inventory Liquidity and Fulfillment Guarantees. *Journal of Business Logistics*, *25*(2), 139–169. https://doi.org/10.1002/j.2158-1592.2004.tb00185.x

Rabinovich, E., & Bailey, J. P. (2004). Physical distribution service quality in Internet retailing: service pricing, transaction attributes, and firm attributes. *Journal of Operations Management*, *21*(6), 651–672. https://doi.org/https://doi.org/10.1016/J.JOM.2003.11.004

Rabinovich, E., & Evers, P. T. (2003). Product Fulfillment in Supply Chains Supporting Internet-Retailing Operations. *Journal of Business Logistics*, *24*(2), 205–236.

Rabinovich, E., Knemeyer, A. M., & Mayer, C. M. (2007). Why do Internet commerce firms incorporate logistics service providers in their distribution channels? The role of transaction costs and network strength. *Journal of Operations Management*, *25*(3), 661–681. https://doi.org/10.1016/j.jom.2006.05.012

Rabinovich, E., Rungtusanatham, M., & Laseter, T. M. (2008). Physical distribution service performance and Internet retailer margins: The drop-shipping context. *Journal of Operations*

Management, *26*(6), 767–780. https://doi.org/10.1016/j.jom.2008.05.001

Rao, S., Goldsby, T. J., Griffis, S. E., & Iyengar, D. (2011). Electronic Logistics Service Quality (e-LSQ): Its Impact on the Customer ' s Purchase Satisfaction and Retention. *Journal of Business Logistics*, *32*(2), 167–179. https://doi.org/https://doi.org/10.1111/j.2158-1592.2011.01014.x

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

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

Retail TouchPoints, Oracle, & Convey. (2018). *Are you winning the last mile of retail?* https://www.oracle.com/webfolder/s/adv/doc1/RT153-SURV-LastMile-Mar-2018.pdf

Rose, W. J., Bell, J. E., & Griffis, S. E. (2022). Inductive research in last-mile delivery routing: Introducing the Re-Gifting heuristic. *Journal of Business Logistics*, *Forthcoming*. https://doi.org/https://doi.org/10.1111/jbl.12318

Rubin, M., Paolini, S., & Crisp, R. J. (2010). A processing fluency explanation of bias against migrants. *Journal of Experimental Social Psychology*, *46*(1), 21–28. https://doi.org/10.1016/j.jesp.2009.09.006

Rungtusanatham, M., Miller, J. W., & Boyer, K. K. (2014). Theorizing, testing, and concluding for mediation in SCM research: Tutorial and procedural recommendations. *Journal of Operations Management*, *32*(3), 99–113. https://doi.org/https://doi.org/10.1016/j.jom.2014.01.002

Rungtusanatham, M., Wallin, C., & Eckerd, S. (2011). The vignette in a scenario-based role-playing experiment. *Journal of Supply Chain Management*, *47*(3), 9–16. https://doi.org/https://doi.org/10.1111/j.1745-493X.2011.03232.x

Russo, I., Masorgo, N., & Gligor, D. M. (2022). Examining the impact of service recovery resilience in the context of product replacement: the roles of perceived procedural and interactional justice. *International Journal of Physical Distribution & Logistics Management*, *52*(8), 638–672. https://doi.org/10.1108/IJPDLM-07-2021-0301

Salari, N., Liu, S., & Shen, Z.-J. M. (2022). Real-Time Delivery Time Forecasting and Promising in Online Retailing: When Will Your Package Arrive? *Manufacturing & Service Operations Management*, *24*(3), 1421–1436. https://doi.org/10.1287/msom.2022.1081

Schneider, C., Weinmann, M., Mohr, P. N. C., & vom Brocke, J. (2021). When the stars shine too bright: The influence of multidimensional ratings on online consumer ratings. *Management Science*, *67*(6), 3871–3898. https://doi.org/10.1287/mnsc.2020.3654

Schwieterman, M. A., Goldsby, T. J., & Croxton, K. L. (2018). Customer and Supplier Portfolios: Can Credit Risks be Managed Through Supply Chain Relationships? *Journal of Business Logistics*, *39*(2), 123–137. https://doi.org/10.1111/jbl.12179

Seiders, K., Voss, G. B., Grewal, D., & Godfrey, A. L. (2005). Do satisfied customers buy more? Examining moderating influences in a retailing context. *Journal of Marketing*, *69*(4), 26–43. https://doi.org/https://doi.org/10.1509/jmkg.2005.69.4.26

Senot, C. (2019). Continuity of Care and Risk of Readmission: An Investigation into the Healthcare Journey of Heart Failure Patients. *Production and Operations Management*, *28*(8), 2008–2030. https://doi.org/10.1111/poms.13027

Serkan Akturk, M., Ketzenberg, M., & Heim, G. R. (2018). Assessing impacts of introducing ship-to-store service on sales and returns in omnichannel retailing: A data analytics study. *Journal of Operations Management*, *61*(1), 15–45. https://doi.org/10.1016/j.jom.2018.06.004

Serra-Garcia, M., & Szech, N. (2023). Incentives and Defaults Can Increase COVID-19 Vaccine Intentions and Test Demand. *Management Science*, *69*(2), 1037–1049. https://doi.org/10.1287/mnsc.2022.4405

Shang, G., Pekgün, P., Ferguson, M., & Galbreth, M. (2017). How much do online consumers really value free product returns? Evidence from eBay. *Journal of Operations Management*, *53–56*, 45–62. https://doi.org/10.1016/j.jom.2017.07.001

Shang, W., & Liu, L. (2011). Promised delivery time and capacity games in time-based competition. *Management Science*, *57*(3), 599–610. https://doi.org/10.1287/mnsc.1100.1292

Simmering, M. J., Fuller, C. M., Richardson, H. A., Ocal, Y., & Atinc, G. M. (2015). Marker Variable Choice, Reporting, and Interpretation in the Detection of Common Method Variance: A Review and Demonstration. *Organizational Research Methods*, *18*(3), 473–511. https://doi.org/https://doi.org/10.1177/1094428114560023

Singh, J., Hansen, M. T., & Podolny, J. M. (2010). The world is not small for everyone: Inequity in searching for knowledge in organizations. *Management Science*, *56*(9), 1415–1438. https://doi.org/10.1287/mnsc.1100.1201

Singh, S., Kemerer, C. F., & Ramasubbu, N. (2021). Innovation in complex assembled electronic products: An analysis of the evolution of television components. *Journal of Operations Management*, *67*(6), 680–703. https://doi.org/10.1002/joom.1139

Sirdeshmukh, D., Singh, J., & Sabol, B. (2002). Consumer Trust , Value , and Loyalty. *Journal of Marketing*, *66*(1), 15–37. https://doi.org/https://doi.org/10.1509/jmkg.66.1.15.18449

Smith, A. K., & Bolton, R. N. (2002). The effect of customers' emotional responses to service failures on their recovery effort evaluations and satisfaction judgments. *Journal of the Academy of Marketing Science*, *30*(1), 5–23. https://doi.org/10.1177/03079450094298

Smith, A. K., Bolton, R. N., & Wagner, J. (1999). A Model of Customer Satisfaction with Service Encounters Involving Failure and Recovery. *Journal of Marketing Research*, *36*(3), 356–372. https://doi.org/10.1177/002224379903600305

Song, H., Tucker, A. L., Graue, R., Moravick, S., & Yang, J. J. (2020). Capacity pooling in hospitals: The hidden consequences of off-service placement. *Management Science*, *66*(9), 3825–3842. https://doi.org/10.1287/mnsc.2019.3395

Staiger, D., & Stock, J. H. (1997). Instrumental Variables Regression with Weak Instruments.

Econometrica, *65*(3), 557. https://doi.org/10.2307/2171753

Stank, T. P., Goldsby, T. J., Vickery, S. K., & Katrina, S. (2003). Logistics Service Performance: Estimating Its Influence On Market Share. *Journal of Business Logistics*, *24*(1), 27–55. https://doi.org/https://doi.org/10.1002/j.2158-1592.2003.tb00031.x

Stewart, D. M., & Chase, R. B. (1999). The impact of human error on delivering service quality. *Production and Operations Management*, *8*(3), 240–263. https://doi.org/10.1111/j.1937-5956.1999.tb00307.x

Stroh, A. M., Erera, A. L., & Toriello, A. (2022). Tactical Design of Same-Day Delivery Systems. *Management Science*, *68*(5), 3444–3463. https://doi.org/10.1287/mnsc.2021.4041

Ta, H., Esper, T. L., & Rossiter, A. (2018). Designing crowdsourced delivery systems : The effect of driver disclosure and ethnic similarity. *Journal of Operations Management*, *60*(1), 19–33. https://doi.org/https://doi.org/10.1016/j.jom.2018.06.001

Ta, H., Esper, T. L., Rossiter Hofer, A., & Sodero, A. (2023). Crowdsourced delivery and customer assessments of e-Logistics Service Quality: An appraisal theory perspective. *Journal of Business Logistics*, *Forthcoming*. https://doi.org/https://doi.org/10.1111/jbl.12327

Taken Smith, K. (2012). Longitudinal study of digital marketing strategies targeting Millennials. *Journal of Consumer Marketing*, *29*(2), 86–92. https://doi.org/10.1108/07363761211206339

Tereyagoglu, N., Peter, F., & Senthil, V. (2018). Multiattribute loss aversion and reference dependence: Evidence from the performing arts industry. *Management Science*, *64*(1), 421–436. https://doi.org/10.1287/mnsc.2016.2605

Thirumalai, S., & Sinha, K. K. (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. https://doi.org/https://doi.org/10.1016/j.jom.2004.10.015

Thirumalai, S., & Sinha, K. K. (2009). Customization strategies in electronic retailing: Implications of customer purchase behavior. *Decision Sciences*, *40*(1), 5–36. https://doi.org/10.1111/j.1540-5915.2008.00222.x

Thirumalai, S., & Sinha, K. K. (2011). Customization of the online purchase process in electronic retailing and customer satisfaction: An online field study. *Journal of Operations Management*, *29*(5), 477–487. https://doi.org/https://doi.org/10.1016/j.jom.2010.11.009

Thomas, R. W., Ueltschy Murfield, M. L., & Ellram, L. M. (2022). Leveraging sustainable supply chain information to alter last-mile delivery consumption: A social exchange perspective. *Sustainable Production and Consumption*, *34*, 285–299. https://doi.org/10.1016/j.spc.2022.09.014

Thürer, M., Fernandes, N. O., & Stevenson, M. (2020). Material Flow Control in High-Variety Make-to-Order Shops: Combining COBACABANA and POLCA. *Production and Operations Management*, *29*(9), 2138–2152. https://doi.org/10.1111/poms.13218

Thürer, M., Land, M. J., Stevenson, M., Fredendall, L. D., & Godinho Filho, M. (2015). Concerning Workload Control and Order Release: The Pre-Shop Pool Sequencing Decision. *Production and Operations Management*, *24*(7), 1179–1192.

https://doi.org/10.1111/poms.12304

Tokar, T., Williams, B. D., & Fugate, B. S. (2020). I Heart Logistics—Just Don't Ask Me to Pay For It: Online Shopper Behavior in Response to a Delivery Carrier Upgrade and Subsequent Shipping Charge Increase. *Journal of Business Logistics*, *41*(3), 182–205. https://doi.org/https://doi.org/10.1111/jbl.12239

Tripathy, M., Bai, J., & Heese, H. S. (2022). Driver collusion in ride-hailing platforms. *Decision Sciences*, *March 2021*, 1–13. https://doi.org/10.1111/deci.12561

U.S. Bureau of Labor Statistics. (2022). *Consumer expenditures 2021*. U.S. Bureau of Labor Statistics. https://www.bls.gov/news.release/cesan.nr0.htm

Umashankar, N., Ward, M. K., & Dahl, D. W. (2017). The benefit of becoming friends: Complaining after service failures leads customers with strong ties to increase loyalty. *Journal of Marketing*, *81*(6), 79–98. https://doi.org/10.1509/jm.16.0125

Venkatesh, V., Bala, H., & Sykes, T. A. (2010). Impacts of information and communication technology implementations on employees' jobs in service organizations in India: A multi-method longitudinal field study. *Production and Operations Management*, *19*(5), 591–613. https://doi.org/10.1111/j.1937-5956.2010.01148.x

Vincenty, T. (1975). Direct and Inverse Solutions of Geodesics on the Ellipsoid with Application of Nested Equations. *Survey Review*, *23*(176), 88–93. https://doi.org/10.1179/sre.1975.23.176.88

Wagner, L., Pinto, C., & Amorim, P. (2021). On the Value of Subscription Models for Online Grocery Retail. *European Journal of Operational Research*, *294*(3), 874–894. https://doi.org/10.1016/j.ejor.2020.05.011

Walmart.com. (2023). *Spark Driver Service Level Standards - My Metrics*. https://drive4spark.walmart.com/My Metrics

Wan, X., Huang, X., & Dong, Y. (2016). The Moderating Role of Product Categories in the Relationship between Online Fulfillment, Procurement, and Consumer Repurchase Intention: A Hierarchical Analysis. *Journal of Supply Chain Management*, *52*(4), 63–76. https://doi.org/10.1111/jscm.12114

Wang, L., Rabinovich, E., & Guda, H. (2022). An analysis of operating efficiency and policy implications in last-mile transportation following Amazon's integration. *Journal of Operations Management*, 1–27. https://doi.org/https://doi.org/10.1002/joom.1172

Wang, L., Rabinovich, E., & Richards, T. J. (2022). Scalability in Platforms for Local Groceries: An Examination of Indirect Network Economies. *Production and Operations Management*, *31*(1), 318–340. https://doi.org/10.1111/poms.13536

Watson, L., & Spence, M. T. (2007). Causes and consequences of emotions on consumer behaviour: A review and integrative cognitive appraisal theory. *European Journal of Marketing*, *41*(5–6), 487–511. https://doi.org/https://doi.org/10.1108/03090560710737570

Williams, L. J., Hartman, N., & Cavazotte, F. (2010). Method variance and marker variables: A review and comprehensive cfa marker technique. *Organizational Research Methods*, *13*(3),

477–514. https://doi.org/10.1177/1094428110366036

Wooldridge, J. M. (2010). *Econometric Analysis of Cross Section and Panel Data*. MIT press. https://doi.org/10.1515/humr.2003.021

Wooldridge, J. M. (2016). *Introductory econometrics: A modern approach*. Nelsen Education.

Xing, Y., & Grant, D. B. (2006). Developing a framework for measuring physical distribution service quality of multi-channel and "pure player" internet retailers. *International Journal of Retail and Distribution Management*, *34*(4–5), 278–289. https://doi.org/10.1108/09590550610660233

Xing, Y., Grant, D. B., McKinnon, A. C., & Fernie, J. (2010). Physical distribution service quality in online retailing. *International Journal of Physical Distribution & Logistics Management*, *40*(5), 415–432. https://doi.org/10.1108/09600031011052859

Zeithaml, V. A., Berry, L. L., & Parasuraman, A. (1996). The behavioral consequences of service quality. *Journal of Marketing*, *60*(2), 31–46. https://doi.org/https://doi.org/10.2307/1251929

Zhan, Y., Zhang, Z., & Ge, J. (2023). Value of Exclusive Doorstep Delivery in the Last-100-Meter Distribution. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.4352455

Zimmermann, M., Sonthalia, B., Oca, A., Martinez-Simon, A., Acar, K., & Sun, Y. (2020). 31st Annual State of Logistics Report. In *Council of supply chain management professionals*.

Zohar, E., Mandelbaum, A., & Shimkin, N. (2002). Adaptive behavior of impatient customers in tele-queues: Theory and empirical support. *Management Science*, *48*(4), 566–583. https://doi.org/10.1287/mnsc.48.4.566.211

V. Conclusion

This dissertation investigates the growing phenomenon of crowdshipping delivery, which has impacted the last mile delivery industry by providing an alternative outsourcing solution to face the rising challenges related to delivery service supply scarcity, uncertainty in driver-customer interaction, and increasing customer expectations (Benjaafar et al., 2019; Daugherty, Bolumole, & Grawe, 2019; Ta et al., 2023). Specifically, this dissertation is an effort to look into the final stage of the order fulfillment process in the crowdshipping context, which is often preferred to third-party logistics outsourcing for hybrid delivery systems by major retailer (Safdar, 2017; Castillo et al., 2022). We investigate the research questions of this dissertation through a multi-method design approach, complementing a rich archival dataset comprised of several million orders retrieved from a Fortune 100 retail crowdshipping platform, with publicly available archival data, and scenario-based experiments.

Essay 1 addresses the first challenge by investigating delivery service providers supply uncertainty. Specifically, essay 1 studies the impact of delivery task remuneration and operational characteristics on drivers' pre-task (Acceptance response time), task (Service time), and post-task behaviors (Driver retention). We found that monetary incentives are not the sole factor influencing drivers' behaviors. Drivers also consider the operational characteristics of the task when accepting, performing, and evaluating a delivery. In the on-demand context of crowshipping delivery, where every minute counts, tasks with greater delivery density can reduce the acceptance response time up to 5 minutes, and the service time by an additional 7 minutes. Considering that service time (i.e., how long it takes for a driver to complete the delivery task) averages 15 minutes, a denser task brings two advantages. On the one hand, the platform could count on an additional driver available for deliveries, which results in, for example, offering an extra time slot to customers. On

the other hand, the platform faces supply scarcity as not requesting an extra driver to be available in the system. Conversely, drivers seem to prefer standard to expedited deliveries, with chances that an expedited delivery would take 4.5 minutes longer to be accepted. Considering the nature of an expedited delivery, which requires an extra time effort from fulfillment operations, 4.5 minutes can severely impact the platform performance and customer outcomes. Finally, supply uncertainty is also driven by the dramatic survival rate of delivery drivers. Monetary incentives, in this case, play the major role to driver retention, with an increase of chances to retain the driver growing from 15% to 35% given an increase of $100 of total remuneration on a given day (i.e., $100 corresponds to 10 completed deliveries), though a larger compensation reduces the frequency with which drivers perform deliveries.

The second study examines a driver's learning experience relative to a delivery task and the context where the delivery took place. Results show the positive impact of driver familiarity on delivery time performance, with a 14% decrease in tardiness for every additional visit of a driver to a customer, and a saving of 2 minutes of tardiness after the third visit. We also investigated the moderation roles of repeating a delivery type and the context of delivery. We found that repeating a delivery type significantly improves performance and potentially reduces tardiness by 1.5 minutes (below, we explain why 1 minute of tardiness matters). However, the effect of repeating the delivery type vanishes as drivers gain more familiarity with the customer. This is especially important for unattended deliveries, which are more difficult to perform for delivery drivers. Finally, despite urban deliveries presenting worse delivery time performance, the interplay with familiarity does not seem to significantly impact the performance. That is, the delivery context does not affect how familiarity improves delivery time performance.

Finally, Essay 3 focuses on how delivery performance (lateness and earliness) shape customers' experience and repurchase behavior with the retailer, and examines important contingency factors in these relationships, namely delivery window length and expedited delivery. Findings from econometric analysis and a scenario-based experiment confirm that late deliveries decrease customer outcomes, with an increase in chances to lose a customer by 2.4% every 1 minute of lateness from the end of the delivery window, but informs that early deliveries also present worse customer outcomes in terms of satisfaction and repurchase behavior, while it decreases the chances to lose a customer by 2.3% for every 1 minute of earliness. This study also show the trade-off between offering a two-hour window vs an eight-hour window. The former increases customer expectations that are more difficult to meet but improves the service experience. The latter decreases expectations but dissatisfies customers, who would consistently provide a 1-star rating when facing a delivery failure. Finally, the experiment further informs on the key role of trust and commitment in the customer-retailer relationship. Delivery failures erode trust regardless of prior expectations. Conversely, customers rata higher levels of commitment when an early delivery occurs with longer delivery windows as compared to a shorter one. Overall, this study informs theory and retailers on the importance of managing delivery window slots and delivery performance, as an effort to effectively handle consumers' time-related expectations on the last mile delivery service.

This research endeavor offers some overarching theoretical contributions for supply chain and operations management literature. First, as supply chains develop and embrace new sharing economy business models, relationship ties become of key importance for a successful delivery provision. Specifically, retaining drivers allows to improve service provision, and retaining customers increases delivery volumes, which in turn affects driver retention and potentially

generates driver-customer familiarity. In crowdshipping, this *spiral* mutually nourishing relationships effect refers to the network effect, which requires, to start, a larger number of drivers than customers (Apte & Davis, 2019; Cullen & Farronato, 2021; Mittal et al., 2021). The conclusion is that retailers and platforms should first invest in expanding the delivery provider supply capacity, but targeting drivers who deliver consistently and are loyal.

Second, as the operations management literature on crowdshipping concentrate their efforts to improve delivery system design and fulfillment operations performance (Benjaafar & Hu, 2020; Mao et al., 2022), we inform this literature focusing on the importance, in this context, of time constraints. Specifically, we emphasize the importance of every minutes. For example, a 12 minutes saving in delivery operations determines a 28.8% increases chances of a customer repurchase a 1.5 minutes of tardiness saved due to driver-customer familiarity (recall the network effect), result in the chance to retain the customer by 2.4%.

As we transition to the overarching managerial contributions of this dissertation, we reconcile here the trade-off between the two-hour vs. eight-hour window with the results of the other two studies. An eight-hour window ensures greater flexibility and efficiency in the design of the last mile delivery service. Hence, platforms have a chance to increase density by consolidating deliveries in fewer delivery tasks to attract drivers and improve delivery performance. However, customers favorably perceive a two-hour delivery window instead. A solution resides in segmenting customers and delivery service. For example, a viable solution is limiting the two-hour slots only to the most profitable customers or premium customers could significantly improve performance, knowing that such a promise requires a platform's extra effort in recruiting delivery drivers.

This dissertation expands the knowledge on last mile delivery and crowdshipping. The purpose of this research endeavor is to unveil the complexities residing in new sharing economy business models for last mile delivery, and address them adopting a supply chain and operations management perspective. We recommend managers to identify and retain those drivers who perform a consistent amount of deliveries and regularly delivers for the platform, to segment customers and assign a premium service level to the most profitable customers, and to focus on the relationship development with the final customer.

Reference list

Apte, U. M., & Davis, M. M. 2019. Sharing economy services: Business model generation. *California Management Review*, 61(2): 104–131.

Benjaafar, S., & Hu, M. 2020. Operations management in the age of the sharing economy: What is old and what is new? *Manufacturing and Service Operations Management*, 22(1): 93–101.

Benjaafar, S., Kong, G., Li, X., & Courcoubetis, C. 2019. Peer-to-peer product sharing: Implications for ownership, usage, and social welfare in the sharing economy. *Management Science*, 65(2): 477–493.

Castillo, V. E., Bell, J. E., Mollenkopf, D. A., & Stank, T. P. 2022. Hybrid last mile delivery fleets with crowdsourcing: A systems view of managing the cost-service trade-off. *Journal of Business Logistics*, 43(1): 36–61.

Cullen, Z., & Farronato, C. 2021. Outsourcing tasks online: Matching supply and demand on peer-to-peer internet platforms. *Management Science*, 67(7): 3985–4003.

Daugherty, P. J., Bolumole, Y., & Grawe, S. J. 2019. The new age of customer impatience: An agenda for reawakening logistics customer service research. *International Journal of Physical Distribution & Logistics Management*, 49(1): 4–32.

Mao, W., Ming, L., Rong, Y., Tang, C. S., & Zheng, H. 2022. On-Demand Meal Delivery Platforms: Operational Level Data and Research Opportunities. *Manufacturing & Service Operations Management*, (Forthcoming).

Mittal, A., Oran Gibson, N., Krejci, C. C., & Marusak, A. A. 2021. Crowd-shipping for urban food rescue logistics. *International Journal of Physical Distribution and Logistics Management*, 51(5): 486–507.

Safdar, K. 2017. Target to Buy Grocery Delivery Startup Shipt for $550 Million. *Wall Street Journal*. Retrieved from https://www.wsj.com/articles/target-to-buy-grocery-delivery-startup-shipt-for-550-million-1513181845

Ta, H., Esper, T. L., Rossiter Hofer, A., & Sodero, A. 2023. Crowdsourced delivery and customer assessments of e-Logistics Service Quality: An appraisal theory perspective. *Journal of Business Logistics*, (Forthcoming).

VI. Appendices

Appendix A – Essay 1 Robustness check and post-hoc analysis

Robustness checks – Acceptance response time

We performed a series of sensitivity analyses to ensure the robustness of our estimations. First, we ensured that the method adopted to address endogeneity for sample bias did not affect the estimations. Hence, we computed the inverse mills ratio replacing delivery zones with zip codes fixed effects as predictors in the first stage, with the intention to capture at more granular level potential differences occurring even among zip codes within the same delivery zones. Results reported in Table 33 for Remuneration provide consistent estimations.

Table 33 – Robustness check results of Table 7 with imr computed with zip code fixed effects

	(1)		(2)		(3)	
	ln (A_i)		ln (A_i)		ln (A_i)	
	β	se	β	se	β	se
cum_{fd}	-0.080****	(0.002)	-0.081****	(0.002)	-0.081****	(0.002)
exp_d	-0.000	(0.000)	0.000	(0.000)	0.000	(0.000)
$dist_i$	-0.000	(0.009)	0.000	(0.009)	0.000	(0.009)
age_d	-0.005****	(0.000)	-0.005****	(0.000)	-0.005****	(0.000)
ndr_z	-0.006****	(0.001)	-0.004****	(0.001)	-0.004****	(0.001)
nor_i	0.217****	(0.005)	0.367****	(0.006)	0.367****	(0.006)
nun_i	0.053****	(0.003)	0.000	(0.003)	-0.000	(0.003)
imr	-1.287****	(0.019)	-1.373****	(0.020)	-1.374****	(0.020)
Store fe	YES		YES		YES	
Month fe	YES		YES		YES	
Day of the week fe	YES		YES		YES	
Time of the day fe	YES		YES		YES	
Remuneration	-0.028****	(0.000)	-0.029****	(0.000)	-0.030****	(0.000)
Density			-0.196****	(0.003)	-0.205****	(0.003)
Expedited			0.258****	(0.007)	0.294****	(0.008)
Remuneration x Density					-0.005****	(0.000)
Remuneration x Expedited					0.010****	(0.001)
Constant	1.602****	(0.029)	1.483****	(0.028)	1.483****	(0.028)
r2	0.080		0.085		0.085	
r2 adjusted	0.080		0.085		0.085	
N	1,667,993		1,667,993		1,667,993	
F	1693.816****		1641.915****		1516.479****	

Note: A_i is Acceptance response time (minutes). $^{+}$ $p < 0.1$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.005$, **** $p < 0.001$.
Reported robust standard errors are clustered on driver ID.

Second, we ensured that the estimation model was not a source of endogeneity. Hence, we adopted a Poisson regression to test the set of hypotheses on A_i (i.e., not ln transformed). Results reported in Table 34 for Remuneration provide consistent estimations.

Table 34 – Robustness check results of Table 7 with Poisson regression

	(1)		(2)		(3)	
	A_i		A_i		A_i	
	β	se	β	se	β	se
cum_{fd}	-0.025****	(0.002)	-0.024****	(0.002)	-0.024****	(0.002)
exp_d	0.002****	(0.000)	0.002****	(0.000)	0.002****	(0.000)
$dist_i$	-0.000	(0.008)	-0.001	(0.008)	-0.001	(0.008)
age_d	-0.005****	(0.000)	-0.005****	(0.000)	-0.005****	(0.000)
ndr_z	-0.006****	(0.000)	-0.005****	(0.000)	-0.005****	(0.000)
nor_i	0.269****	(0.004)	0.283****	(0.005)	0.282****	(0.005)
nun_i	0.012****	(0.003)	-0.022****	(0.003)	-0.021****	(0.003)
imr	-1.026****	(0.016)	-1.037****	(0.017)	-1.039****	(0.017)
Store fe	YES		YES		YES	
Month fe	YES		YES		YES	
Day of the week fe	YES		YES		YES	
Time of the day fe	YES		YES		YES	
Remuneration	-0.014****	(0.000)	-0.015****	(0.000)	-0.017****	(0.000)
Density			-0.156****	(0.003)	-0.165****	(0.003)
Expedited			-0.160****	(0.006)	-0.093****	(0.007)
Remuneration x Density					-0.005****	(0.000)
Remuneration x Expedited					0.019****	(0.001)
Constant	2.788****	(0.022)	2.834****	(0.022)	2.834****	(0.022)
χ^2	41290.110		44620.495		45708.295	
N	1,667,993		1,667,993		1,667,993	

Note: A_i is Acceptance response time (minutes). $^{+}$ $p < 0.1$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.005$, **** $p < 0.001$. Reported robust standard errors are clustered on driver ID.

Robustness checks – Service time

We performed the same estimations but including the inverse mills ratio computed with zip code fixed effects. Results on Table 35 confirm the same estimations.

Table 35 – Robustness check results of Table 7 with imr computed with zip code fixed effects

	(1)		(2)		(3)	
	ST_i		ST_i		ST_i	
	β	se	β	se	β	se
cum_{fd}	-0.228****	(0.006)	-0.219****	(0.005)	-0.220****	(0.005)
exp_d	-0.015****	(0.001)	-0.011****	(0.001)	-0.011****	(0.001)
$dist_i$	0.104***	(0.034)	0.089***	(0.031)	0.089***	(0.031)
age_d	0.022****	(0.002)	0.020****	(0.001)	0.020****	(0.001)
ndr_z	-0.011****	(0.001)	0.019****	(0.001)	0.019****	(0.001)
nor_i	0.501****	(0.017)	1.482****	(0.021)	1.481****	(0.021)
nun_i	5.724****	(0.017)	4.831****	(0.016)	4.808****	(0.016)
Imr zip	-0.630****	(0.069)	-1.200****	(0.066)	-1.211****	(0.066)
Store fe	YES		YES		YES	
Month fe	YES		YES		YES	
Day of the week fe	YES		YES		YES	
Time of the day fe	YES		YES		YES	
Remuneration	0.073****	(0.001)	0.048****	(0.001)	0.043****	(0.001)
Density			-3.699****	(0.012)	-3.785****	(0.013)
Expedited			-1.244****	(0.022)	-1.202****	(0.025)
Remuneration x Density					-0.052****	(0.002)
Remuneration x Expedited					0.006^{+}	(0.003)
Constant	6.114****	(0.094)	6.488****	(0.087)	6.507****	(0.087)
r2	0.203		0.305		0.306	
r2 adjusted	0.203		0.305		0.306	
N	1,667,993		1,667,993		1,667,993	
F	8575.988****		13163.886****		12387.741****	

Note: ST_i is Service time (minutes). $^{+}$ $p < 0.1$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.005$, **** $p < 0.001$. Reported robust standard errors are clustered on driver ID.

Robustness checks – Survival analysis

We performed the same estimations with different estimation models. Specifically, we performed the same set of estimation analyses but utilized a fully parametric proportional hazard regression model, which allowed us to specify the distribution of the hazard rate (Bradburn et al., 2003). Following the literature (Bhattacharjee et al., 2007; Azadegan, Patel, & Parida, 2013; Tereyağoğlu, Fader, & Veeraraghavan, 2017; De Vries, Roy, & De Koster, 2018; Batt et al., 2019; Li et al., 2022), we specify a Weibull distribution for the hazard rate. Table 36 reports the estimations when specifying a Weibull distribution, which are consistent with the results of the main analysis.

Table 36 – Cox proportional hazard regression model specifying a Weibull distribution for remuneration per day

	(1) $DRAT_{dy}$ β	se	(2) $DRAT_{dy}$ β	se
exp_{dy}			0.005****	(0.000)
$ntask_{dy}$			-0.159****	(0.006)
ndr_{y}			0.001****	(0.000)
Month fe			YES	
Day of the week fe			YES	
Time of the day fe			YES	
Remuneration day	-0.005****	(0.000)	-0.000	(0.000)
Constant	-2.087****	(0.008)	-2.941****	(0.019)
χ^2	4850.891		70054.69	
N	862,199		862,199	
N clusters	77,561		77,561	

Note: $DRAT_{dy}$ is driver attrition. $^{+}$ $p < 0.1$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.005$, **** $p < 0.001$. Reported robust standard errors are clustered on driver ID.

In addition, we performed the same set of estimation analyses but utilizing Aalen's additive hazard model (Aalen, 1989), which allows more flexibility than a proportional hazard model when incorporating time-dependent covariates (Bradburn et al., 2003). Prior literature found this methodology useful in investigating supply chain and operations management research questions (Ramasubbu & Kemerer, 2016; Mao et al., 2019). The estimations in Table 37 provide consistent estimations: Remuneration day has a negative effect on the odds of a driver disappearing from the dataset.

Table 37 – Aalen's additive hazard model estimations for remuneration per day

	(1) $DRAT_{dy}$ β	z	p	(2) $DRAT_{dy}$ β	z	p
exp_{dy}				-0.00100	56.547	0.000
$ntask_{dy}$				-0.01332	-21.182	0.000
ndr_{y}				0.00008	9.143	0.000
Month fe				YES		
Day of the week fe				YES		
Time of the day fe				YES		
Remuneration day	-0.00110	-81.762	0.000	-0.00001	-6.005	0.000
Constant	0.13875	209.716	0.000	0.08333	65.828	0.000
McFadden's pseudo R^2	0.966			0.999		
N	862,199			862,199		

Note: $DRAT_{dy}$ is driver attrition. $^{+}$ $p < 0.1$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.005$, **** $p < 0.001$. Reported robust standard errors are clustered on driver ID.

Appendix B – Essay 2 Robustness check and post-hoc analysis

Robustness checks

To assess the robustness of these estimations, we performed a series of robustness checks. First, we addressed endogeneity relative to the estimation model, by performing the same analysis in Table 15 but replacing the dependent variable Tardiness with a binary variable Delivery Performance, taking 1 if the order was late, 0 otherwise, and adopting a Logit Regression. Estimations reported in Table 38 show consistent results.

Next, we performed the same analysis of Table 15, but including an additional control variable (Restart), which is an indicator taking 1 if order *i* was the first in a given day for the driver. Prior literature discusses the importance of the *warm-up* effect that a worker may suffer on the first task of the day (Ibanez et al., 2018). Estimations included in Table 39 report similar results.

Table 38 – Robustness check results of Table 4 – Estimations with Logit Regression and Binary Dependent Variable Delivery Performance (1 = late; 0 otherwise)

	(1) Del Perf		(2) Del Perf		(3) Del Perf	
	β	se	β	se	β	se
Distance from store	0.221****	(0.001)	0.222****	(0.001)	0.222****	0.221****
Driver age	-0.005****	(0.000)	-0.005****	(0.000)	-0.005****	-0.005****
Customer subscription	-0.632****	(0.004)	-0.631****	(0.004)	-0.630****	-0.632****
Order size	-0.024****	(0.000)	-0.024****	(0.000)	-0.024****	-0.024****
Expedited	-2.285****	(0.016)	-2.287****	(0.016)	-2.287****	-2.285****
Driver fatigue	0.010****	(0.002)	0.010****	(0.002)	0.010****	0.010****
Store fe	YES		YES		YES	
Month fe	YES		YES		YES	
Delivery day of the week fe	YES		YES		YES	
Delivery hour fe	YES		YES		YES	
Familiarity	-0.175****	(0.006)	-0.127****	(0.006)	-0.150****	(0.009)
Delivery type repetition			-0.183****	(0.011)	-0.283****	(0.015)
Urban			0.055****	(0.007)	0.054****	(0.007)
Repeat unattended x Familiarity					0.110****	(0.011)
Urban x Familiarity					-0.026*	(0.010)
Constant	-0.678****	(0.164)	-0.708****	(0.164)	-0.712****	(0.164)
χ^2	183866.669		184992.035		185100.633	
Pseudo-R2	0.230		0.230		0.230	
N	6,963,868		6,963,868		6,963,868	

Note: $^{+}$ $p < 0.1$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.005$, **** $p < 0.001$; Reported robust standard errors are clustered on driver ID.

Table 39 – Robustness check results of Table 4 – Poisson Regression Estimations including Restart

	(1) Tardiness β	se	(2) Tardiness β	se	(3) Tardiness β	se
Distance from store	0.214****	(0.001)	0.214****	(0.001)	0.214****	(0.001)
Driver age	-0.001****	(0.000)	-0.001****	(0.000)	-0.001****	(0.000)
Customer subscription	-0.382****	(0.003)	-0.382****	(0.003)	-0.381****	(0.003)
Order size	-0.023****	(0.000)	-0.023****	(0.000)	-0.023****	(0.000)
Expedited	-1.755****	(0.018)	-1.756****	(0.018)	-1.756****	(0.018)
Driver fatigue	0.004****	(0.001)	0.004****	(0.001)	0.004****	(0.001)
Restart	-0.049****	(0.005)	-0.049****	(0.005)	-0.049****	(0.005)
Store fe	YES		YES		YES	
Month fe	YES		YES		YES	
Delivery day of the week fe	YES		YES		YES	
Delivery hour fe	YES		YES		YES	
Familiarity	-0.152****	(0.005)	-0.111****	(0.005)	-0.120****	(0.008)
Delivery type repetition			-0.158****	(0.010)	-0.220****	(0.015)
Urban			0.042****	(0.005)	0.041****	(0.005)
Repeat unattended x Familiarity					0.065****	(0.011)
Urban x Familiarity					-0.019*	(0.009)
Constant	3.484****	(0.127)	3.461****	(0.126)	3.459****	(0.126)
χ^2	337614.456		339198.301		339342.155	
Pseudo-R2	0.297		0.297		0.298	
N	6,963,868		6,963,868		6,963,868	

Note: $^{+}$ $p < 0.1$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.005$, **** $p < 0.001$; Reported robust standard errors are clustered on driver ID.

Post-hoc analysis

We performed a post-hoc analysis to explore which delivery type between unattended and attended worsen the delivery performance. Hence, we assessed whether the characteristics of the task (attended vs unattended) present intrinsic difficulties for delivery drivers. We performed the estimations of Table 15 but replacing Repeat unattended with a binary variable with value 1 if the delivery of order i is unattended (i.e., we do not account for a driver’s prior experience in completing an unattended delivery). Interestingly, estimations reported in Table 40 show that unattended deliveries a more challenging to perform.

Table 40 – Robustness check results of Table 4 – Poisson Regression Estimations with moderators Binary Variables Unattended and Urban

	(1) Tardiness β	se	(2) Tardiness β	se	(3) Tardiness β	se
Distance from store	0.215****	(0.001)	0.214****	(0.001)	0.214****	(0.001)
Driver age	-0.001****	(0.000)	-0.001****	(0.000)	-0.001****	(0.000)
Customer subscription	-0.383****	(0.002)	-0.379****	(0.002)	-0.379****	(0.002)
Order size	-0.023****	(0.000)	-0.023****	(0.000)	-0.023****	(0.000)
Expedited	-1.762****	(0.013)	-1.755****	(0.013)	-1.755****	(0.013)
Driver fatigue	0.005****	(0.001)	0.005****	(0.001)	0.005****	(0.001)
Store fe	YES		YES		YES	
Month fe	YES		YES		YES	
Delivery day of the week fe	YES		YES		YES	
Delivery hour fe	YES		YES		YES	
Familiarity	-0.152****	(0.004)	-0.155****	(0.004)	-0.241****	(0.013)
Unattended			0.165****	(0.004)	0.172****	(0.004)
Urban			0.043****	(0.005)	0.042****	(0.005)
Unattended x Familiarity					0.105****	(0.012)
Urban x Familiarity					-0.019*	(0.008)
Constant	3.458****	(0.114)	3.278****	(0.114)	3.272****	(0.114)
χ^2	413167.736		416030.523		416409.643	
Pseudo-R2	0.297		0.298		0.298	
N	13,927,531		13,927,531		13,927,531	

Note: $^{+}$ $p < 0.1$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.005$, **** $p < 0.001$; Reported robust standard errors are clustered on driver ID.

Appendix C – Essay 3: Study 1 sensitivity analysis

Endogeneity for sample bias

Despite our effort to ensure the validity of the measures and dataset, we investigated whether the presence of the customer rating in the dataset was not affected by endogeneity for sample bias, which could affect the analysis. Hence, following recommended practices to address sample selection bias (Lu et al., 2018), we consider whether the estimated effects of such *smaller* sample could systematically bias the estimations due to unobservable factors. First, following recent operations management literature (Alexander, Boone, & Lynn, 2021), we used a logit regression utilizing the full sample (~14 million orders) to determine the effect of the key variables of interest and control variables on the likelihood that a customer provides the rating. Table 41 (Model 1) reports the results. We found that the characteristics of the delivery, of the customer, and the delivery performance may affect whether the customer rated the delivery.

Table 41 – Results estimation models to address endogeneity for sample bias

	(1) Rating provided order i		(2) Rating provided order i	
	β	se	β	se
order size	0.004^{****}	(0.000)		
distance from store	-0.039^{****}	(0.001)		
n late previous orders	-0.035^{****}	(0.001)		
n early previous orders	-0.053^{****}	(0.001)		
subscription	0.126^{****}	(0.003)		
unattended	-0.506^{****}	(0.003)		
Lateness	-0.009^{****}	(0.000)		
Earliness	0.002^{****}	(0.000)		
Window length	-1.663^{****}	(0.002)		
Expedited	0.285^{****}	(0.004)		
Rating provided order i-1			1.047^{****}	(0.001)
Intercept	-0.192^{**}	(0.071)	-1.400^{****}	(0.037)
Month fe	YES		YES	
Day of the week fe	YES		YES	
Time of the day fe	YES		YES	
χ^2	1185888.377		741217.422	
McFadden's pseudo R^2	0.125		0.081	
N	14,251,769		14,251,769	

Note: $^{+}$ $p < 0.1$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.005$, **** $p < 0.001$. Robust standard errors reported in parenthesis.

In light of these results, to avoid this source of endogeneity, we adopt the Heckman's sample selection model (Heckman, 1979). We computed the Inverse-Mills ratio, and included it as a control variables. We estimated the Inverse-Mills ratio through a probit regression predicting the occurrence of observing customer rating in the dataset (Gambeta, Koka, & Hoskisson, 2019). Thus, we calculated the probability that a customer rated the delivery experience of order i utilizing the full sample (~14 million orders) including $PrevRating_{i-1c}$, a predictor not included in the main regression model to meet the exclusion restriction (Shang et al., 2017). We performed the probit regression on a binary dependent variable (Rating available 1 = yes; 0 otherwise) with the following selection equation:

$$(26)\ \ Rating_i = \begin{cases} 1 & if\ Z_i\Gamma + \varepsilon_i > 0, \\ 0 & \quad otherwise \end{cases}$$

Where Z_i is the vector of independent variables, Γ is the vector of coefficients, and ε_i is the error term. $Z_i\Gamma$ is specified as follows:

$$(27)\ \ Z_i = \gamma_0 + \gamma_1 PrevRating_{i-1c} + \gamma_{2im} month_{im} + \gamma_3 dow_{iw} + \gamma_{4it} time_{it}$$

Where $PrevRating_{i-1c}$ a binary variable taking 1 if the customer provided a rating on his/her previous order, $month_{im}$, dow_{iw}, $time_{it}$, are the categorical dummy variables for, respectively, month m, day of the week w, and time of the day t. Table 41 (Model 2) reports the result of equation (6), and shows that customers who rated the previous order are more likely to rate current order (β = 1.047, SE = 0.001). From equation (6), we obtain the inverse Mills ratio (IMR) and include it in all regression models in this study.

We performed the same estimation models used in the main analysis including the inverse mills ratio as a control variable, and found consistent estimations (Table 42 and Table 43). Thus, we conclude that endogeneity for sample bias is not of concern in this study.

Table 42 – Estimation results for Customer Satisfaction including Inverse Mills Ratio as control variable

	(1)		(2)					
	Ln(rating)		Ln(rating)		Ln(rating)		Ln(rating)	
Order size	0.000****	(0.000)	0.000****	(0.000)	0.000	(0.000)	0.000****	(0.000)
Distance from store	-0.000	(0.000)	-0.001	(0.000)	-0.011****	(0.002)	-0.002****	(0.000)
N late prev orders	-0.002****	(0.000)	-0.002****	(0.000)	-0.012****	(0.002)	-0.001****	(0.000)
N early prev orders	-0.001	(0.001)	-0.000	(0.001)	-0.015****	(0.004)	0.002***	(0.001)
Subscription	0.020****	(0.001)	0.019****	(0.001)	0.022****	(0.001)	0.021****	(0.001)
Unattended	0.002	(0.001)	0.001	(0.001)	0.003****	(0.001)	0.001	(0.001)
Imr	-0.058****	(0.001)	-0.058****	(0.001)	-0.058****	(0.001)	-0.058****	(0.001)
Month fe	YES		YES		YES		YES	
Day of the week fe	YES		YES		YES		YES	
Clock hour fe	YES		YES		YES		YES	
Store fe	YES		YES		YES		YES	
Lateness	-0.039****	(0.001)	-0.040****	(0.001)	-0.115****	(0.018)	-0.040****	(0.001)
Earliness	-0.022****	(0.001)	-0.023****	(0.001)	-0.077****	(0.012)	-0.024****	(0.001)
Window length			-0.110****	(0.008)	-0.283****	(0.042)	-0.470****	(0.056)
Expedited			-0.034****	(0.002)	-0.062****	(0.007)	0.212****	(0.036)
Lateness x Win len					0.022****	(0.005)		
Earliness x Win len					0.019****	(0.004)		
Lateness x Exp							-0.089****	(0.013)
Earliness x Exp							-0.085****	(0.014)
_cons	1.686****	(0.028)	1.801****	(0.030)	1.786****	(0.040)	2.173****	(0.065)
F	321.949****		304.699****		289.077****		289.076****	
R^2	0.008		0.008		0.008		0.008	
Adjusted R^2	0.008		0.008		0.008		0.008	
N	1,394,399		1,394,399		1,394,399		1,394,399	
rmse	0.379		0.379		0.379		0.379	

Note: robust standard errors reported in parenthesis. $^{+}$ $p < 0.1$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.005$, **** $p < 0.001$

Table 43 – Estimation results for Repurchase Intentions including Inverse Mills Ratio as control variable

	(1)		(2)		(3)		(4)	
	Repurchase		Repurchase		Repurchase		Repurchase	
Order size	0.009****	(0.000)	0.009****	(0.000)	0.009****	(0.000)	0.009****	(0.000)
Distance from store	0.018****	(0.001)	0.018****	(0.001)	0.014*	(0.006)	0.017****	(0.001)
N late prev orders	0.008****	(0.001)	0.008****	(0.001)	0.002	(0.006)	0.008****	(0.001)
N early prev orders	0.006**	(0.002)	0.007***	(0.002)	-0.007	(0.012)	0.009****	(0.002)
Subscription	0.243****	(0.004)	0.242****	(0.004)	0.244****	(0.004)	0.243****	(0.004)
Unattended	0.038****	(0.003)	0.037****	(0.003)	0.038****	(0.003)	0.037****	(0.003)
Imr	-0.205****	(0.002)	-0.205****	(0.002)	-0.205****	(0.002)	-0.205****	(0.002)
Month fe	YES		YES		YES		YES	
Day of the week fe	YES		YES		YES		YES	
Clock hour fe	YES		YES		YES		YES	
Store fe	YES		YES		YES		YES	
Lateness	-0.078****	(0.005)	-0.079****	(0.005)	-0.137**	(0.052)	-0.081****	(0.005)
Earliness	-0.017****	(0.003)	-0.017****	(0.003)	-0.037	(0.033)	-0.018****	(0.003)
Window length			-0.528****	(0.032)	-0.620****	(0.115)	-0.772****	(0.145)
Expedited			0.063****	(0.007)	0.044*	(0.020)	0.224*	(0.094)
Lateness x Win len					0.019	(0.016)		
Earliness x Win len					0.006	(0.012)		
Lateness x Exp							-0.044	(0.036)
Earliness x Exp							-0.068*	(0.034)
_cons	3.176****	(0.061)	3.711****	(0.071)	3.622****	(0.102)	3.971****	(0.164)
χ^2	80585.745		85181.935		85190.151		85195.410	
pseudo R^2	0.096		0.096		0.096		0.096	
N	1,394,399		1,394,399		1,394,399		1,394,399	

Note: robust standard errors reported in parenthesis. $^{+}$ $p < 0.1$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.005$, **** $p < 0.001$

Robustness checks

We performed a series of sensitivity analysis to ensure the robustness of our results. First, we ensured that the choice of the estimation model did not affect the estimations relative to customer rating. Thus, following recent literature (Perdikaki, Peng, & Heim, 2015; Akturk, Mallipeddi, & Jia, 2022), we used an Ordered Logit Regression model to estimate the impact of predictors on an rating. Hence, we treat customer rating as an ordered discrete variable (i.e., not continuous) (Chang, Dasgupta, & Hilary, 2010; Singh, Hansen, & Podolny, 2010; Schwieterman, Goldsby, & Croxton, 2018). We tested this model as follows:

$$(28)\ Pr(Rating_i = k) = Pr(\alpha_{k-1} < y_i^* < \alpha_k)$$

Where $k \in [1, 5]$, $\alpha_{k-1}\ \alpha_k$ are the thresholds, and y_i^* is the latent variable specified as follow:

$$(29)\ y_i^* = \alpha_0 + \alpha_1 \widehat{Lateness}_i + \alpha_2 \widehat{Earliness}_i + \mathrm{BX}_i + \varepsilon_i$$

Results in Table 44 show consistent estimations, suggesting that the choice of the estimation model did not affect the results.

Table 44 – Estimation results for Customer Satisfaction using Ordered Logit Regression

	(1) Rating		(2) Rating		(3) Rating		(4) Rating	
Order size	0.003^{****}	(0.000)	0.003^{****}	(0.000)	0.002^{****}	(0.000)	0.003^{****}	(0.000)
Distance from store	0.006^{*}	(0.003)	0.004	(0.003)	-0.043^{*}	(0.017)	-0.003	(0.003)
N late prev orders	-0.010^{****}	(0.002)	-0.011^{****}	(0.002)	-0.057^{****}	(0.016)	-0.007^{**}	(0.002)
N early prev orders	-0.001	(0.004)	0.001	(0.004)	-0.066^{*}	(0.028)	0.015^{***}	(0.005)
Subscription	0.145^{****}	(0.007)	0.139^{****}	(0.007)	0.152^{****}	(0.008)	0.146^{****}	(0.007)
Unattended	0.050^{****}	(0.007)	0.048^{****}	(0.007)	0.057^{****}	(0.008)	0.047^{****}	(0.007)
Imr	YES		YES		YES		YES	
Month fe	YES		YES		YES		YES	
Day of the week fe	YES		YES		YES		YES	
Clock hour fe	YES		YES		YES		YES	
Store fe	-0.297^{****}	(0.011)	-0.300^{****}	(0.011)	-0.642^{****}	(0.127)	-0.302^{****}	(0.011)
Lateness	-0.146^{****}	(0.008)	-0.148^{****}	(0.008)	-0.395^{****}	(0.087)	-0.156^{****}	(0.009)
Earliness			-0.647^{****}	(0.064)	-1.417^{****}	(0.294)	-2.413^{****}	(0.400)
Window length			-0.188^{****}	(0.018)	-0.315^{****}	(0.048)	1.026^{****}	(0.261)
Expedited					0.098^{*}	(0.039)		
Lateness x Win len					0.087^{**}	(0.031)		
Earliness x Win len							-0.446^{****}	(0.096)
Lateness x Exp							-0.411^{****}	(0.097)
χ^2	5957.501		5993.982		5994.285		5994.140	
pseudo R^2								
N	1,394,399		1,394,399		1,394,399		1,394,399	

Note: Robust standard errors reported in parenthesis. $^{+}$ $p < 0.1$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.005$, **** $p < 0.001$

Second, upon consulting prior literature (Rao et al., 2011), we verified whether the estimations for repurchase intentions were not affected by customer satisfaction, which is typically an important predictor for future purchase decisions. Table 45 reports the same estimations models used for repurchase intentions but including customer rating as additional control variable. We found consistent estimation, excluding that customer satisfaction affects these estimations.

Table 45 – Estimation results for Repurchase Intentions including Customer Satisfaction as control variable

	(1) Repurchase		(2) Repurchase		(3) Repurchase		(4) Repurchase	
Order size	0.009^{****}	(0.000)	0.009^{****}	(0.000)	0.008^{****}	(0.000)	0.009^{****}	(0.000)
Distance from store	0.019^{****}	(0.001)	0.019^{****}	(0.001)	0.014^{*}	(0.006)	0.017^{****}	(0.001)
N late prev orders	0.010^{****}	(0.001)	0.010^{****}	(0.001)	0.004	(0.006)	0.011^{****}	(0.001)
N early prev orders	0.005^{*}	(0.002)	0.006^{***}	(0.002)	-0.008	(0.012)	0.008^{****}	(0.002)
Subscription	0.255^{****}	(0.004)	0.254^{****}	(0.004)	0.257^{****}	(0.004)	0.255^{****}	(0.004)
Unattended	0.034^{****}	(0.003)	0.033^{****}	(0.003)	0.034^{****}	(0.003)	0.033^{****}	(0.003)
Imr	0.001	(0.001)	0.001	(0.001)	0.001	(0.001)	0.001	(0.001)
Month fe	YES		YES		YES		YES	
Day of the week fe	YES		YES		YES		YES	
Clock hour fe	YES		YES		YES		YES	
Store fe	YES		YES		YES		YES	
Lateness	-0.085^{****}	(0.005)	-0.086^{****}	(0.005)	-0.147^{***}	(0.052)	-0.089^{****}	(0.005)
Earliness	-0.019^{****}	(0.003)	-0.019^{****}	(0.003)	-0.042	(0.033)	-0.021^{****}	(0.003)
Window length			-0.534^{****}	(0.031)	-0.633^{****}	(0.115)	-0.808^{****}	(0.145)
Expedited			0.051^{****}	(0.007)	0.031	(0.020)	0.232^{*}	(0.094)
Lateness x Win len					0.019	(0.016)		
Earliness x Win len					0.007	(0.012)		
Lateness x Exp							-0.049	(0.036)
Order size							-0.076^{*}	(0.034)
_cons	2.846^{****}	(0.061)	3.387^{****}	(0.071)	3.303^{****}	(0.102)	3.679^{****}	(0.164)
χ^2	73611.008		78094.169		78101.520		78106.404	
pseudo R^2	0.092		0.092		0.092		0.092	
N	1,394,399		1,394,399		1,394,399		1,394,399	

Note: robust standard errors reported in parenthesis. $^{+}$ $p < 0.1$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.005$, **** $p < 0.001$

Finally, we conducted an additional robustness check for the effects on repurchase intentions by adding zip code fixed effects to the regression models. Recent literature (Lim, Gao, & Tan, 2020) suggests adding zip codes fixed effects to control for heterogeneity factors that affect customers' purchase decisions relative to a specific geographical area (Lim et al., 2020; Lim, Wang, & Webster, 2023). Estimations reported in Table 46 confirm the estimations found in the main analysis.

Table 46 – Estimation results for Repurchase Intentions including zip code fixed effects

	(1) Repurchase		(2) Repurchase		(3) Repurchase		(4) Repurchase	
Order size	0.009****	(0.001)	0.008****	(0.000)	0.008****	(0.001)	0.008****	(0.001)
Distance from store	0.001	(0.011)	0.001	(0.011)	-0.012	(0.022)	-0.003	(0.013)
N late prev orders	0.000	(0.010)	0.000	(0.010)	-0.016	(0.020)	0.002	(0.010)
N early prev orders	0.006	(0.016)	0.007	(0.017)	-0.027	(0.014)	0.013	(0.020)
Subscription	0.255****	(0.004)	0.251****	(0.004)	0.256****	(0.004)	0.254****	(0.004)
Unattended	0.024****	(0.006)	0.022****	(0.007)	0.026****	(0.006)	0.022***	(0.007)
Imr	YES		YES		YES		YES	
Month fe	YES		YES		YES		YES	
Day of the week fe	YES		YES		YES		YES	
Clock hour fe	YES		YES		YES		YES	
Store fe	YES		YES		YES		YES	
Lateness	-0.170*	(0.083)	-0.168*	(0.084)	-0.319	(0.174)	-0.170*	(0.084)
Earliness	-0.095*	(0.042)	-0.093*	(0.043)	-0.160	(0.105)	-0.098*	(0.045)
Window length			-0.849****	(0.197)	-1.107***	(0.392)	-1.667*	(0.673)
Expedited			-0.034	(0.064)	-0.086	(0.094)	0.525	(0.288)
Lateness x Win len					0.048	(0.030)		
Earliness x Win len					0.021	(0.023)		
Lateness x Exp							-0.194	(0.120)
Order size							-0.197	(0.117)
_cons	2.597****	(0.346)	3.451****	(0.517)	3.271****	(0.483)	4.306****	(0.974)
χ^2	186081.934		189208.699		189214.250		189213.755	
pseudo R^2	0.104		0.104		0.104		0.104	
N	1,394,399		1,394,399		1,394,399		1,394,399	

Note: robust standard errors reported in parenthesis. $^{+}$ $p < 0.1$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.005$, **** $p < 0.001$

Appendix D – Essay 3: Study 2 experimental setup and vignette

*Instructions to participants***:** In this task you will be provided with an online purchasing scenario. After the scenario, you will be asked several questions about your purchasing preferences, followed by some demographic questions. Please read the questions carefully, and take as much time as needed. There are no right or wrong answers.

Common module: You need to order a week's worth of groceries. You place your order online with Grocer.com. Grocer.com is a grocery retailer that provides delivery services for online shopping.

Experimental cues module delivery window length:

Two hours	Five hours	Eight hours
Upon finalizing the order, you chose the two-hour delivery window that was most convenient for you.	Upon finalizing the order, you chose the five-hour delivery window that was most convenient for you.	Upon finalizing the order, you chose the eight-hour delivery window that was most convenient for you.

Experimental cues module delivery performance: After two days, you receive the following email from Grocer.com:

Example delivery confirmation email for the 10:00AM-12:00PM delivery window and late delivery performance

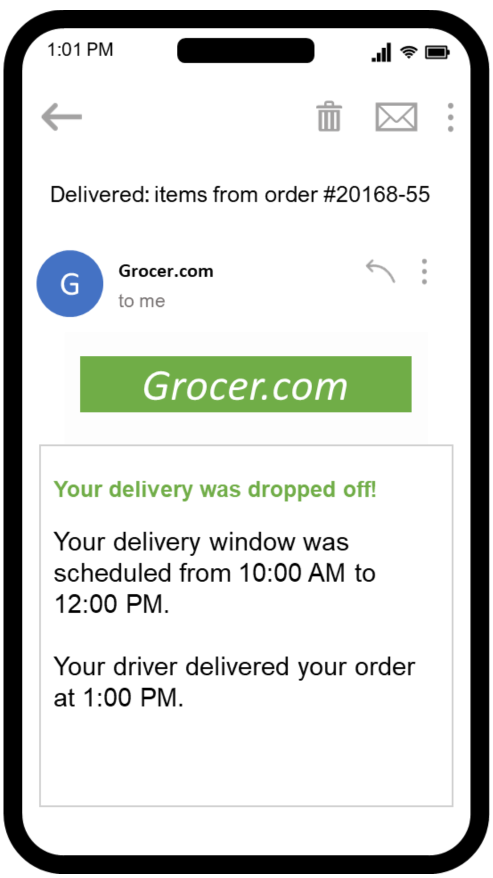


Table 47 – Experimental cues module delivery performance - Message in the email screenshot

Window length		**Late delivery**	**Early delivery**	**On-time delivery**
Two hours	10AM 12PM	Your delivery window was scheduled from 10:00 AM to 12:00 PM. Your driver delivered your order at 1:00 PM.	Your delivery window was scheduled from 10:00AM to 12:00 PM. Your driver delivered your order at 9:00 AM.	Your delivery window was scheduled from 10:00AM to 12:00 PM. Your driver delivered your order at 10:15 AM.
	12PM 2 PM	Your delivery window was scheduled from 12:00 PM to 2:00 PM. Your driver delivered your order at 3:00 PM.	Your delivery window was scheduled from 12:00 PM to 2:00 PM. Your driver delivered your order at 11:00 AM.	Your delivery window was scheduled from 12:00 PM to 2:00 PM. Your driver delivered your order at 12:15 PM.
	2 PM 4 PM	Your delivery window was scheduled from 2:00 PM to 4:00 PM. Your driver delivered your order at 5:00 PM.	Your delivery window was scheduled from 2:00 PM to 4:00 PM. Your driver delivered your order at 1:00 PM.	Your delivery window was scheduled from 2:00 PM to 4:00 PM. Your driver delivered your order at 2:15 PM.
	4 PM 6 PM	Your delivery window was scheduled from 4:00 PM to 6:00 PM. Your driver delivered your order at 7:00 PM.	Your delivery window was scheduled from 4:00 PM to 6:00 PM. Your driver delivered your order at 3:00 PM.	Your delivery window was scheduled from 4:00 PM to 6:00 PM. Your driver delivered your order at 4:15 PM.
	6 PM 8PM	Your delivery window was scheduled from 6:00 PM to 8:00 PM. Your driver delivered your order at 9:00 PM.	Your delivery window was scheduled from 6:00 PM to 8:00 PM. Your driver delivered your order at 5:00 PM.	Your delivery window was scheduled from 6:00 PM to 8:00 PM. Your driver delivered your order at 6:15 PM.
Five hours	10AM 3 PM	Your delivery window was scheduled from 10:00AM to 3:00 PM. Your driver delivered your order at 1:00 PM.	Your delivery window was scheduled from 10:00AM to 3:00 PM. Your driver delivered your order at 1:00 PM.	Your delivery window was scheduled from 10:00AM to 3:00 PM. Your driver delivered your order at 10:15 AM
	1 PM 6 PM	Your delivery window was scheduled from 1:00 PM to 6:00 PM. Your driver delivered your order at 7:00 PM.	Your delivery window was scheduled from 1:00 PM to 6:00 PM. Your driver delivered your order at 12:00 AM.	Your delivery window was scheduled from 1:00 PM to 6:00 PM. Your driver delivered your order at 1:15 PM.
	3 PM 8 PM	Your delivery window was scheduled from 3:00 PM to 8:00 PM. Your driver delivered your order at 9:00 PM.	Your delivery window was scheduled from 3:00 PM to 8:00 PM. Your driver delivered your order at 2:00 PM.	Your delivery window was scheduled from 3:00 PM to 8:00 PM. Your driver delivered your order at 3:15 PM.
Eight hours	10AM 6 PM	Your delivery window was scheduled from 10:00AM to 6:00 PM. Your driver delivered your order at 7:00 PM.	Your delivery window was scheduled from 10:00AM to 6:00 PM. Your driver delivered your order at 9:00 AM.	Your delivery window was scheduled from 10:00AM to 6:00 PM. Your driver delivered your order at 10:15 PM.
	12PM 8 PM	Your delivery window was scheduled from 12:00 PM to 8:00 PM. Your driver delivered your order at 9:00 PM.	Your delivery window was scheduled from 12:00 PM to 8:00 PM. Your driver delivered your order at 11:00AM.	Your delivery window was scheduled from 12:00 PM to 8:00 PM. Your driver delivered your order at 12:15 PM.

Appendix F – Essay 3: Study 2 sensitivity analysis

Robustness checks

We ensured the robustness of our estimations through a series of analysis. First, we performed the main analysis without control variables to exclude failed randomization (Lonati et al., 2018). Results in Tables 48-49-50-51 confirm the estimations obtained in the main analysis.

Table 48 – Mediation results of Process Model 4 without control variables

Condition	Focal lever	Mediation paths effects $\boldsymbol{X \xrightarrow{a} M \xrightarrow{b} DV}$	Indirect effects [95% CI] $\boldsymbol{\theta = a \times b}$	Direct effect [CI] $\boldsymbol{X \xrightarrow{C} DV}$
On-time vs Late	Late	$L \xrightarrow{-1.819} T \xrightarrow{0.248} SAT$	-0.451 [-0.681; -0.266]	$L \xrightarrow{-1.466} SAT$
		$L \xrightarrow{-1.869} C \xrightarrow{0.430} SAT$	-0.804 [-1.071; -0.553]	
		$L \xrightarrow{-1.819} T \xrightarrow{0.194} REP$	-0.403 [-0.625; -0.223]	$L \xrightarrow{-0.954} REP$
		$L \xrightarrow{-1.869} C \xrightarrow{0.505} REP$	-1.138 [-1.427; -0.863]	
On-time vs Early	Early	$E \xrightarrow{-1.118} T \xrightarrow{0.312} SAT$	-0.348 [-0.522; -0.213]	$E \xrightarrow{-0.743} SAT$
		$E \xrightarrow{-1.061} C \xrightarrow{0.553} SAT$	-0.587 [-0.813; -0.381]	
		$E \xrightarrow{-1.118} T \xrightarrow{0.290} REP$	-0.324 [-0.500; -0.195]	$E \xrightarrow{-0.372} REP$
		$E \xrightarrow{-1.061} C \xrightarrow{0.686} REP$	-0.728 [-0.990; -0.488]	
Early vs Late	Late	$E \xrightarrow{-0.701} T \xrightarrow{0.367} SAT$	-0.257 [-0.427; -0.122]	$L \xrightarrow{-0.328} SAT$
		$E \xrightarrow{-0.807} C \xrightarrow{0.567} SAT$	-0.458 [-0.660; -0.274]	
		$E \xrightarrow{-0.701} T \xrightarrow{0.377} REP$	-0.264 [-0.429; -0.130]	$L \xrightarrow{-0.265} REP$
		$E \xrightarrow{-0.807} C \xrightarrow{0.672} REP$	-0.542 [-0.768; -0.330]	

Note: Reported effects are all significant at $p < 0.05$. Results with 5,000 bootstraps

Table 49 – Moderated Mediation results of On-time vs Late without control variables

On-time vs Late	(1) **Trust**	(2) **Commitment**	(3) **SAT**	(4) **REP**
Constant	5.381*** (0.174)	4.94*** (0.156)	2.249*** (0.252)	1.522*** (0.275)
Late	-1.92*** (0.247)	-2.155*** (0.222)	-1.466*** (0.134)	-0.954*** (0.147)
Trust			0.248*** (0.043)	0.222*** (0.046)
Commitment			0.431*** (0.047)	0.609*** (0.051)
WL 1 (2h vs 5h)	-0.27 (0.244)	-0.203 (0.22)		
WL 2 (2h vs 8h)	-0.417* (0.249)	-0.6*** (0.224)		
Late x WL 1	0.069 (0.347)	0.214 (0.312)		
Late x WL 2	0.244 (0.353)	0.667** (0.317)		
F-value (*df*)	33.09*** (5,306)	43.727*** (5,306)	317.509*** (3,308)	262.442*** (3,308)
R^2	0.592	0.646	0.869	0.848
IMM indirect effect via Trust		IMM WL 1	0.017 (0.085) CI [-0.156,0.182]	0.015 (0.078) CI [-0.146,0.166]
		IMM WL 2	0.060 (0.089) CI [-0.121,0.235]	0.054 (0.083) CI [-0.097,0.230]
Indirect effect via Trust		IE WL 2 hours	-0.476 (0.117) CI [-0.726,-0.275]	-0.426 (0.118) CI [-0.681,-0.217]
		IE WL 5 hours	-0.459 (0.119) CI [-0.720,-0.260]	-0.410 (0.119) CI [-0.677,-0.207]
		IE WL 8 hours	-0.415 (0.114) CI [-0.665,-0.224]	-0.372 (0.105) CI [-0.597,-0.190]
IMM indirect effect via Commitment		IMM WL 1	0.092 (0.128) CI [-0.157,0.344]	0.131 (0.178) CI [-0.224,0.475]
		IMM WL 2	0.287 (0.131) CI [0.034,0.553]	0.406 (0.191) CI [0.042,0.793]
Indirect effect via Commitment		IE WL 2 hours	-0.928 (0.145) CI [-1.22,-0.652]	-1.313 (0.166) CI [-1.644,-0.998]
		IE WL 5 hours	-0.835 (0.164) CI [-1.175,-0.535]	-1.182 (0.194) CI [-1.571,-0.811]
		IE WL 8 hours	-0.641 (0.148) CI [-0.956,-0.375]	-0.906 (0.174) CI [-1.256,-0.576]

Note: First stage common to both SAT and REP presented in models 1 and 2. WL1 compares 2h to 5h, and WL2 compares 2h to 8h. * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$. Standard error in parenthesis. Abbreviations: IMM = Index of Moderated Mediation; IE = Indirect effect; CI = Confidence Interval.

Table 50 – Moderated Mediation results of On-time vs Early without control variables

On-time vs Early	(1) **Trust**	(2) **Commitment**	(3) **SAT**	(4) **REP**
Constant	5.381*** (0.196)	4.94*** (0.189)	1.348*** (0.232)	0.812*** (0.243)
Early	-1.415*** (0.276)	-1.732*** (0.267)	-0.743*** (0.118)	-0.372*** (0.124)
Trust			0.312*** (0.043)	0.29*** (0.045)
Commitment			0.553*** (0.044)	0.686*** (0.046)
WL 1 (2h vs 5h)	-0.27 (0.276)	-0.203 (0.267)		
WL 2 (2h vs 8h)	-0.417 (0.281)	-0.600** (0.272)		
Early x WL 1	0.216 (0.388)	0.632* (0.375)		
Early x WL 2	0.698* (0.395)	1.424*** (0.383)		
F-value (*df*)	10.544*** (5,311)	12.276*** (5,311)	209.783*** (3,313)	206.597*** (3,313)
R^2	0.381	0.406	0.817	0.815
IMM indirect effect via Trust		IMM WL 1	0.067 (0.118) CI [-0.162,0.304]	0.063 (0.112) CI [-0.156,0.288]
		IMM WL 2	0.217 (0.133) CI [-0.029,0.491]	0.202 (0.126) CI [-0.02,0.473]
Indirect effect via Trust		IE WL 2 hours	-0.441 (0.114) CI [-0.686,-0.242]	-0.41 (0.109) CI [-0.651,-0.221]
		IE WL 5 hours	-0.373 (0.104) CI [-0.601,-0.196]	-0.348 (0.102) CI [-0.57,-0.17]
		IE WL 8 hours	-0.223 (0.105) CI [-0.44,-0.033]	-0.208 (0.095) CI [-0.409,-0.028]
IMM indirect effect via Commitment		IMM WL 1	0.349 (0.196) CI [-0.042,0.734]	0.433 (0.239) CI [-0.044,0.89]
		IMM WL 2	0.787 (0.224) CI [0.36,1.25]	0.976 (0.275) CI [0.45,1.524]
Indirect effect via Commitment		IE WL 2 hours	-0.957 (0.168) CI [-1.304,-0.65]	-1.188 (0.191) CI [-1.569,-0.835]
		IE WL 5 hours	-0.608 (0.155) CI [-0.931,-0.316]	-0.755 (0.19) CI [-1.141,-0.404]
		IE WL 8 hours	-0.171 (0.172) CI [-0.516,0.158]	-0.212 (0.21) CI [-0.639,0.198]

Note: First stage common to both SAT and REP presented in models 1 and 2. WL1 compares 2h to 5h, and WL2 compares 2h to 8h. * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$. Standard error in parenthesis. Abbreviations: IMM = Index of Moderated Mediation; IE = Indirect effect; CI = Confidence Interval.

Table 51 – Moderated Mediation results of On-time vs Early without control variables

Early vs Late	(1) **Trust**	(2) **Commitment**	(3) **SAT**	(4) **REP**
Constant	3.967*** (0.189)	3.207*** (0.176)	0.33 (0.207)	0.136 (0.198)
Late	-0.505* (0.27)	-0.423* (0.252)	-0.329*** (0.123)	-0.265** (0.117)
Trust			0.367*** (0.05)	0.377*** (0.048)
Commitment			0.567*** (0.053)	0.672*** (0.051)
WL 1 (2h vs 5h)	-0.054 (0.267)	0.429* (0.248)		
WL 2 (2h vs 8h)	0.28 (0.272)	0.824*** (0.253)		
Late x WL 1	-0.147 (0.38)	-0.417 (0.354)		
Late x WL 2	-0.454 (0.387)	-0.757** (0.36)		
F-value (*df*)	4.472*** (5,309)	8.262*** (5,309)	144.385*** (3,311)	194.807*** (3,311)
R^2	0.26	0.343	0.763	0.808
IMM indirect effect via Trust		IMM WL 1	-0.054 (0.139) CI [-0.335,0.216]	-0.056 (0.143) CI [-0.349,0.219]
		IMM WL 2	-0.167 (0.153) CI [-0.499,0.119]	-0.171 (0.157) CI [-0.495,0.122]
Indirect effect via Trust		IE WL 2 hours	-0.185 (0.107) CI [-0.407,0.006]	-0.191 (0.11) CI [-0.427,0.012]
		IE WL 5 hours	-0.239 (0.105) CI [-0.466,-0.055]	-0.246 (0.107) CI [-0.481,-0.056]
		IE WL 8 hours	-0.352 (0.13) CI [-0.634,-0.126]	-0.362 (0.131) CI [-0.637,-0.13]
IMM indirect effect via Commitment		IMM WL 1	-0.237 (0.212) CI [-0.664,0.171]	-0.28 (0.235) CI [-0.753,0.163]
		IMM WL 2	-0.429 (0.219) CI [-0.88,-0.018]	-0.508 (0.25) CI [-1.007,-0.027]
Indirect effect via Commitment		IE WL 2 hours	-0.24 (0.148) CI [-0.531,0.048]	-0.284 (0.173) CI [-0.633,0.044]
		IE WL 5 hours	-0.476 (0.157) CI [-0.803,-0.186]	-0.564 (0.171) CI [-0.913,-0.243]
		IE WL 8 hours	-0.669 (0.171) CI [-1.021,-0.351]	-0.792 (0.192) CI [-1.182,-0.425]

Note: First stage common to both SAT and REP presented in models 1 and 2. WL1 compares 2h to 5h, and WL2 compares 2h to 8h. * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$. Standard error in parenthesis. Abbreviations: IMM = Index of Moderated Mediation; IE = Indirect effect; CI = Confidence Interval.

Second, we ensured that social approval and social desirability did significantly affect the estimations of our model and potentially create demand effect (Lonati et al., 2018). We used multiple strategies and common practices to address social desirability bias. First, we ensured participants anonymity to exclude any bias related to participants answering the survey to be favorably viewed (Eckerd et al., 2021). Second, we ensured participants that there were no right or wrong answers (Fisher, 1993). Finally, we collected an additional 4-items 7-point (1 = strongly disagree; 7 = strongly agree) scale ($\alpha = 0.81$) adapted from (Fisher, 1993), to assess the extent to which the participant (1) I sometimes feel resentful when I don't get my way, (2) There have been times when I felt like rebelling against people in authority even though I knew they were right, (3) There have been times when I was quite jealous of the good fortune of others, and (4) I am sometimes irritated by people who ask favors of me. We added this scale as additional control to the main analyses and found consistent estimations. Hence, social desirability is not a concern.

Finally, we performed the analysis including the 77 observations deleted due to failed manipulation check, to avoid any issue related to sample bias. Results reported in Tables 52-53-54-55 provide similar estimations to the main analysis, hence excluding that the 77 removed observations could affect the models.

Table 52 – Mediation results of Process Model 4 with full sample (N = 549)

Condition	**Focal lever**	**Mediation paths effects** $\boldsymbol{X \xrightarrow{a} M \xrightarrow{b} DV}$	**Indirect effects [95% CI]** $\boldsymbol{\theta = a \times b}$	**Direct effect [CI]** $\boldsymbol{X \xrightarrow{c} DV}$
On-time vs Late	Late	$L \xrightarrow{-1.785} T \xrightarrow{0.277} SAT$	-0.495 [-0.713; -0.312]	$L \xrightarrow{-1.339} SAT$
		$L \xrightarrow{-1.816} C \xrightarrow{0.459} SAT$	-0.834 [-1.071; -0.605]	
		$L \xrightarrow{-1.785} T \xrightarrow{0.238} REP$	-0.424 [-0.635; -0.237]	$L \xrightarrow{-0.888} REP$
		$L \xrightarrow{-1.816} C \xrightarrow{0.639} REP$	-1.160 [-1.430; -0.909]	
On-time vs Early	Early	$E \xrightarrow{-1.156} T \xrightarrow{0.364} SAT$	-0.421 [-0.610; -0.273]	$E \xrightarrow{-0.682} SAT$
		$E \xrightarrow{-1.057} C \xrightarrow{0.512} SAT$	-0.542 [-0.750; -0.355]	
		$E \xrightarrow{-1.156} T \xrightarrow{0.340} REP$	-0.392 [-0.571; -0.238]	$E \xrightarrow{-0.352} REP$
		$E \xrightarrow{-1.057} C \xrightarrow{0.637} REP$	-0.674 [-0.926; -0.444]	
Early vs Late	Late	$E \xrightarrow{-0.673} T \xrightarrow{0.391} SAT$	-0.264 [-0.433; -0.128]	$L \xrightarrow{-0.373} SAT$
		$E \xrightarrow{-0.820} C \xrightarrow{0.518} SAT$	-0.425 [-0.625; -0.248]	
		$E \xrightarrow{-0.673} T \xrightarrow{0.406} REP$	-0.273 [-0.451; -0.130]	$L \xrightarrow{-0.302} REP$
		$E \xrightarrow{-0.820} C \xrightarrow{0.621} REP$	-0.509 [-0.729; -0.310]	

Note: Reported effects are all significant at $p < 0.05$. Controls included: Age, Gender, Education, Income, Channel, Time orientation. Results with 5,000 bootstraps

Table 53 – Moderated Mediation results of On-time vs Late with full sample (N = 549)

On-time vs Late	(1) **Trust**	(2) **Commitment**	(3) **SAT**	(4) **REP**
Constant	4.901*** (0.541)	4.276*** (0.511)	2.046*** (0.397)	1.533*** (0.435)
Late	-1.932*** (0.243)	-2.123*** (0.23)	-1.339*** (0.122)	-0.888*** (0.133)
Trust			0.277*** (0.04)	0.238*** (0.044)
Commitment			0.459*** (0.042)	0.639*** (0.046)
WL 1 (2h vs 5h)	-0.283 (0.222)	-0.211 (0.209)		
WL 2 (2h vs 8h)	-0.431* (0.222)	-0.586*** (0.21)		
Late x WL 1	0.142 (0.333)	0.239 (0.315)		
Late x WL 2	0.253 (0.338)	0.63** (0.319)		
Age	0.003 (0.005)	0.005 (0.005)	-0.007** (0.003)	-0.005 (0.004)
Gender	0.135 (0.142)	-0.111 (0.135)	0.019 (0.099)	-0.024 (0.108)
Education	-0.3** (0.152)	-0.211 (0.144)	-0.169 (0.105)	0.114 (0.115)
Income	-0.15 (0.133)	0.023 (0.126)	-0.21** (0.093)	-0.266*** (0.101)
Channel	0.06 (0.133)	0.259** (0.125)	0.092 (0.092)	0.181* (0.101)
Time orientation	0.138 (0.112)	0.135 (0.106)	0.081 (0.077)	-0.021 (0.084)
F-value (*df*)	17.863*** (11,364)	21.302*** (11,364)	130.846*** (9,366)	108.707*** (9,366)
R^2	0.592	0.626	0.873	0.853
IMM indirect effect via Trust		IMM WL 1	0.039 (0.092) CI [-0.143,0.228]	0.034 (0.08) CI [-0.122,0.199]
		IMM WL 2	0.07 (0.098) CI [-0.118,0.277]	0.06 (0.086) CI [-0.097,0.249]
Indirect effect via Trust		IE WL 2 hours	-0.536 (0.122) CI [-0.796,-0.315]	-0.459 (0.121) CI [-0.722,-0.249]
		IE WL 5 hours	-0.496 (0.113) CI [-0.735,-0.296]	-0.425 (0.109) CI [-0.666,-0.233]
		IE WL 8 hours	-0.465 (0.114) CI [-0.709,-0.265]	-0.399 (0.104) CI [-0.619,-0.219]
IMM indirect effect via Commitment		IMM WL 1	0.11 (0.13) CI [-0.155,0.367]	0.153 (0.182) CI [-0.204,0.51]
		IMM WL 2	0.289 (0.138) CI [0.025,0.562]	0.403 (0.191) CI [0.039,0.793]
Indirect effect via Commitment		IE WL 2 hours	-0.975 (0.139) CI [-1.249,-0.715]	-1.357 (0.165) CI [-1.687,-1.046]
		IE WL 5 hours	-0.865 (0.153) CI [-1.179,-0.589]	-1.204 (0.19) CI [-1.607,-0.85]
		IE WL 8 hours	-0.685 (0.138) CI [-0.969,-0.432]	-0.954 (0.165) CI [-1.291,-0.641]

Note: First stage common to both SAT and REP presented in models 1 and 2. WL1 compares 2h to 5h, and WL2 compares 2h to 8h. * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$. Standard error in parenthesis. Abbreviations: IMM = Index of Moderated Mediation; IE = Indirect effect; CI = Confidence Interval.

Table 54 – Moderated Mediation results of On-time vs Early with full sample (N = 549)

On-time vs Early	(1) **Trust**	(2) **Commitment**	(3) **SAT**	(4) **REP**
Constant	5.367*** (0.573)	4.51*** (0.579)	1.772*** (0.42)	0.881* (0.452)
Early	-1.457*** (0.261)	-1.616*** (0.263)	-0.682*** (0.11)	-0.352*** (0.118)
Trust			0.364*** (0.04)	0.34*** (0.043)
Commitment			0.513*** (0.04)	0.637*** (0.043)
WL 1 (2h vs 5h)	-0.288 (0.245)	-0.154 (0.247)		
WL 2 (2h vs 8h)	-0.452* (0.245)	-0.525** (0.247)		
Early x WL 1	0.128 (0.36)	0.508 (0.363)		
Early x WL 2	0.723** (0.362)	1.136*** (0.365)		
Age	0.005 (0.005)	0.006 (0.006)	-0.009** (0.004)	-0.006 (0.004)
Gender	-0.122 (0.152)	-0.028 (0.153)	0.042 (0.105)	0.125 (0.113)
Education	-0.127 (0.162)	0.049 (0.163)	-0.154 (0.111)	0.033 (0.119)
Income	-0.009 (0.145)	0.121 (0.146)	-0.229** (0.1)	-0.098 (0.108)
Channel	0.043 (0.145)	0.365** (0.146)	0.162 (0.101)	0.186* (0.108)
Time orientation	0.005 (0.118)	-0.016 (0.119)	-0.013 (0.082)	-0.001 (0.088)
F-value (*df*)	6.764*** (11,378)	6.825*** (11,378)	86.013*** (9,380)	79.651*** (9,380)
R^2	0.406	0.407	0.819	0.808
IMM indirect effect via Trust		IMM WL 1	0.047 (0.135) CI [-0.216,0.323]	0.044 (0.126) CI [-0.202,0.301]
		IMM WL 2	0.263 (0.146) CI [-0.003,0.571]	0.245 (0.137) CI [-0.021,0.528]
Indirect effect via Trust		IE WL 2 hours	-0.531 (0.13) CI [-0.811,-0.299]	-0.495 (0.127) CI [-0.758,-0.271]
		IE WL 5 hours	-0.484 (0.118) CI [-0.738,-0.279]	-0.451 (0.116) CI [-0.697,-0.248]
		IE WL 8 hours	-0.267 (0.107) CI [-0.489,-0.07]	-0.249 (0.1) CI [-0.465,-0.067]
IMM indirect effect via Commitment		IMM WL 1	0.26 (0.184) CI [-0.102,0.621]	0.324 (0.218) CI [-0.112,0.743]
		IMM WL 2	0.582 (0.197) CI [0.217,0.99]	0.724 (0.242) CI [0.27,1.219]
Indirect effect via Commitment		IE WL 2 hours	-0.828 (0.153) CI [-1.158,-0.545]	-1.03 (0.182) CI [-1.394,-0.685]
		IE WL 5 hours	-0.568 (0.154) CI [-0.887,-0.282]	-0.706 (0.184) CI [-1.089,-0.364]
		IE WL 8 hours	-0.246 (0.146) CI [-0.54,0.036]	-0.306 (0.176) CI [-0.66,0.035]

Note: First stage common to both SAT and REP presented in models 1 and 2. WL1 compares 2h to 5h, and WL2 compares 2h to 8h. * p < 0.1; ** p < 0.05; *** p < 0.01. Standard error in parenthesis. Abbreviations: IMM = Index of Moderated Mediation; IE = Indirect effect; CI = Confidence Interval.

Table 55 – Moderated Mediation results of On-time vs Early with full sample (N = 549)

Early vs Late	(1) **Trust**	(2) **Commitment**	(3) **SAT**	(4) **REP**
Constant	3.965*** (0.578)	3.03*** (0.548)	1.241*** (0.463)	0.879* (0.448)
Late	-0.525* (0.268)	-0.54** (0.255)	-0.373*** (0.125)	-0.302** (0.121)
Trust			0.391*** (0.051)	0.406*** (0.049)
Commitment			0.518*** (0.053)	0.621*** (0.052)
WL 1 (2h vs 5h)	-0.202 (0.26)	0.337 (0.246)		
WL 2 (2h vs 8h)	0.183 (0.264)	0.583** (0.251)		
Late x WL 1	-0.033 (0.377)	-0.338 (0.358)		
Late x WL 2	-0.421 (0.385)	-0.507 (0.366)		
Age	-0.009 (0.006)	0.008 (0.006)	-0.008* (0.005)	-0.002 (0.004)
Gender	-0.046 (0.165)	-0.038 (0.157)	0.03 (0.127)	0.016 (0.123)
Education	0.202 (0.181)	0.195 (0.172)	-0.159 (0.139)	0.01 (0.135)
Income	-0.049 (0.155)	0.006 (0.147)	-0.289** (0.119)	-0.226* (0.115)
Channel	0.259 (0.158)	0.47*** (0.15)	-0.047 (0.121)	0.003 (0.118)
Time orientation	0.048 (0.123)	-0.084 (0.116)	-0.058 (0.093)	-0.124 (0.09)
F-value (*df*)	2.429*** (11,320)	4.514*** (11,320)	46.641*** (9,322)	61.318*** (9,322)
R^2	0.278	0.367	0.752	0.795
IMM indirect effect via Trust		IMM WL 1	-0.013 (0.146) CI [-0.307,0.283]	-0.014 (0.15) CI [-0.304,0.292]
		IMM WL 2	-0.165 (0.162) CI [-0.498,0.124]	-0.171 (0.162) CI [-0.514,0.135]
Indirect effect via Trust		IE WL 2 hours	-0.206 (0.11) CI [-0.437,-0.001]	-0.213 (0.115) CI [-0.458,-0.004]
		IE WL 5 hours	-0.219 (0.109) CI [-0.448,-0.019]	-0.227 (0.111) CI [-0.456,-0.019]
		IE WL 8 hours	-0.37 (0.139) CI [-0.672,-0.128]	-0.384 (0.141) CI [-0.694,-0.139]
IMM indirect effect via Commitment		IMM WL 1	-0.175 (0.192) CI [-0.564,0.202]	-0.21 (0.221) CI [-0.655,0.231]
		IMM WL 2	-0.263 (0.194) CI [-0.673,0.093]	-0.315 (0.228) CI [-0.797,0.111]
Indirect effect via Commitment		IE WL 2 hours	-0.28 (0.141) CI [-0.578,-0.023]	-0.335 (0.164) CI [-0.666,-0.023]
		IE WL 5 hours	-0.455 (0.147) CI [-0.76,-0.19]	-0.545 (0.162) CI [-0.882,-0.244]
		IE WL 8 hours	-0.542 (0.156) CI [-0.873,-0.265]	-0.65 (0.179) CI [-1.022,-0.327]

Note: First stage common to both SAT and REP presented in models 1 and 2. WL1 compares 2h to 5h, and WL2 compares 2h to 8h. * p < 0.1; ** p < 0.05; *** p < 0.01. Standard error in parenthesis. Abbreviations: IMM = Index of Moderated Mediation; IE = Indirect effect; CI = Confidence Interval.

Post-hoc analyses

In the second post-hoc analysis, we verified the non-significant interaction results of the econometric analysis for the effect lateness and earliness on repurchase intentions. We verified that such result hold in the experiment by studying the moderation on the direct effect of lateness (earliness) on repurchase intentions using Process Model 5. The interactions do not show significant coefficients (Table 56). Thus, we confirm the results from the dataset that window length does not moderate the direct effect of lateness and earliness on repurchase intention.

Table 56 – Moderation results of Process Model 5 with moderation in the direct effect of Lateness (L) and Earliness (E) on Repurchase intentions (REP)

	On-time vs Late Repurchase intentions	**On-time vs Early Repurchase intentions**	**Early vs Late Repurchase intentions**
Constant	1.789*** (0.467)	0.969** (0.489)	0.748* (0.443)
Late	-0.947*** (0.215)		-0.192 (0.196)
Early		-0.431** (0.210)	
Trust	0.219*** (0.046)	0.665*** (0.048)	0.678*** (0.053)
Commitment	0.599*** (0.052)	0.299*** (0.046)	0.370*** (0.049)
Window Length 1	0.150 (0.185)	0.169 (0.198)	0.013 (0.193)
Window Length 2	-0.233 (0.189)	-0.186 (0.202)	0.200 (0.199)
Condition x Window Length 1	-0.168 (0.26)	-0.155 (0.279)	0.053 (0.275)
Condition x Window Length 2	0.020 (0.267)	0.365 (0.287)	-0.314 (0.283)
Age	-0.002 (0.004)	-0.006 (0.005)	-0.002 (0.005)
Gender	-0.052 (0.116)	0.083 (0.120)	-0.021 (0.120)
Education	0.167 (0.125)	0.080 (0.129)	0.021 (0.133)
Income	-0.297*** (0.108)	-0.063 (0.115)	-0.238** (0.113)
Channel	0.182* (0.109)	0.138 (0.117)	-0.023 (0.117)
Time orientation	-0.033 (0.089)	-0.004 (0.096)	-0.115 (0.093)
F-value (*df*)	63.94*** (13,298)	48.472*** (13,303)	45.509*** (13,301)
R^2	0.858	0.822	0.814

Note: Results from the second stage of mediation. WL 1 compares 2h to 5h, and WL compares 2h to 8h. * p < 0.1; ** p < 0.05; *** p < 0.01. Standard error in parenthesis.

The last post-hoc analysis refers to additional variables we collected in the survey. Specifically, as a form of commitment, customers typically provide retailers with reviews and feedback that are then used by retailers to improve their performance (Umashankar, Ward, & Dahl, 2017; Akturk et al., 2022). Hence, this post-hoc analysis investigates customers' behavior relative to writing a feedback to the online retailer after experiencing a service failure.

We asked respondents to first write a review relative to their delivery experience. Then, we asked them whether the review they wrote was negative, neutral, or positive. The majority of respondents exposed to the early and late condition indicated their review as negative or neutral, whereas those exposed to the on-time condition indicated their review as a positive one. An Ologit regression with Review mood as dependent variable (coding: 0 = negative review; 1 = neutral review; 2 = positive review), indicates that respondents exposed to the early and late conditions significantly rated their review more as negative and neutral as compared to those exposed to the on-time condition (see Table 57 - Model 1).

In addition, we measured Review length, which capture the extent of complaining in terms of number of words in the review (M = 21.29, SD = 18) (Umashankar et al., 2017). We tested whether customers exposed to a service failure would significantly write a longer review, thus commit extra effort to provide the retailer with a feedback. Results from a Poisson regression (see Table 57 - Model 2) show that those exposed to late and early condition significantly wrote a longer review (see Figure 30).

Table 57 – Results for Post-hoc analysis 2: The effect of Late and Early on review mood and review length

	(1) Review mood		(2) Review length	
Late	-4.573^{***}	(0.329)	0.635^{***}	(0.027)
Early	-2.965^{***}	(0.294)	0.596^{***}	(0.027)
Age	-0.003	(0.008)	0.005^{***}	(0.001)
Gender	-0.338	(0.211)	0.286^{***}	(0.020)
Income	-0.189	(0.223)	-0.080^{***}	(0.021)
Education	0.120	(0.241)	-0.215^{***}	(0.022)
Time orientation	-0.026	(0.167)	-0.081^{***}	(0.016)
Intercept			2.784^{***}	(0.076)
χ^2	298.659		1141.294	
McFadden’s pseudo R^2	0.299		0.141	
N	472		472	

Note: Standard errors in parentheses + p < 0.10, ** p < 0.05, *** p < 0.0

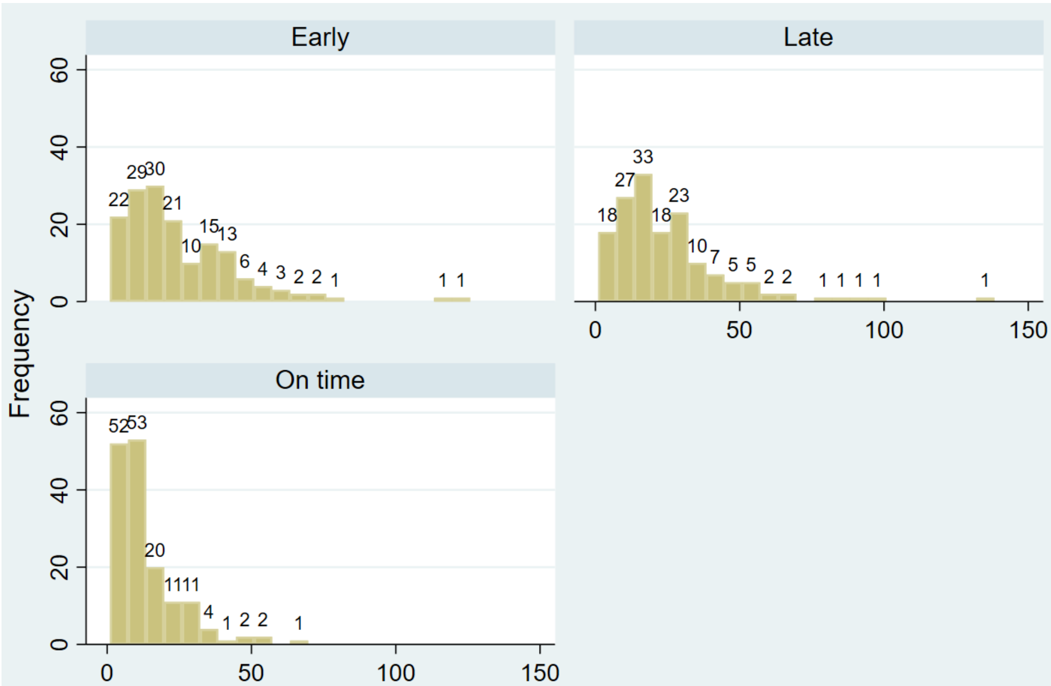


Figure 30 – Distribution of Review length across conditions (Review length: min = 1, max = 132)

Reference list

Aalen, O. O. 1989. A linear regression model for the analysis of life times. *Statistics in Medicine*, 8(8): 907–925.

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

Alexander, D., Boone, C., & Lynn, M. 2021. The effects of tip recommendations on customer tipping, satisfaction, repatronage, and spending. *Management Science*, 67(1): 146–165.

Azadegan, A., Patel, P. C., & Parida, V. 2013. Operational slack and venture survival. *Production and Operations Management*, 22(1): 1–18.

Batt, R. J., Diwas S., K. C., Staats, B. R., & Patterson, B. W. 2019. The Effects of Discrete Work Shifts on a Nonterminating Service System. *Production and Operations Management*, 28(6): 1528–1544.

Bhattacharjee, S., Gopal, R. D., Lertwachara, K., Marsden, J. R., & Telang, R. 2007. The effect of digital sharing technologies on music markets: A survival analysis of albums on ranking charts. *Management Science*, 53(9): 1359–1374.

Bradburn, M. J., Clark, T. G., Love, S. B., & Altman, D. G. 2003. Survival Analysis Part II: Multivariate data analysis- An introduction to concepts and methods. *British Journal of Cancer*, 89(3): 431–436.

Chang, Y. Y., Dasgupta, S., & Hilary, G. 2010. CEO ability, pay, and firm performance. *Management Science*, 56(10): 1633–1652.

De Vries, J., Roy, D., & De Koster, R. 2018. Worth the wait? How restaurant waiting time influences customer behavior and revenue. *Journal of Operations Management*, 63(May): 59–78.

Eckerd, S., DuHadway, S., Bendoly, E., Carter, C., & Kaufmann, L. 2021. On making experimental design choices: Discussions on the use and challenges ofdemand effects, incentives, deception, samples, and vignettes. *Journal of Operations Management*, 67(2): 261–275.

Fisher, R. 1993. Social Desirability Bias and the Validity of Indirect Questioning. *Journal of Consumer Research*, 20(2): 303.

Gambeta, E., Koka, B. R., & Hoskisson, R. E. 2019. Being too good for your own good: A stakeholder perspective on the differential effect of firm-employee relationships on innovation search. *Strategic Management Journal*, 40(1): 108–126.

Heckman, J. J. 1979. Sample Selection Bias as a Specification Error. *Econometrica*, 47(1): 153.

Ibanez, M. R., Clark, J. R., Huckman, R. S., & Staats, B. R. 2018. Discretionary task ordering: Queue management in radiological services. *Management Science*, 64(9): 4389–4407.

Li, Z., Liang, C., Hong, Y., & Zhang, Z. 2022. How Do On-demand Ridesharing Services Affect Traffic Congestion? The Moderating Role of Urban Compactness. *Production and Operations Management*, 31(1): 239–258.

Lim, S. F. W. T., Gao, F., & Tan, T. 2020. Slow and Sure, or Fast and Furious? An Empirical Study About Omnichannel Demand Sensitivity to Fulfillment Lead Time. *SSRN Electronic Journal*, 1–63.

Lim, S. F. W. T., Wang, Q., & Webster, S. 2023. Do it right the first time: Vehicle routing with home delivery attempt predictors. *Production and Operations Management*, Forthcomin.

Lonati, S., Quiroga, B. F., Zehnder, C., & Antonakis, J. 2018. On doing relevant and rigorous experiments: Review and recommendations. *Journal of Operations Management*, 64(1): 19–40.

Lu, G., Ding, X. (David), Peng, D. X., & Hao-Chun Chuang, H. 2018. Addressing endogeneity in operations management research: Recent developments, common problems, and directions for future research. *Journal of Operations Management*, 64(1): 53–64.

Mao, W., Ming, L., Rong, Y., Tang, C. S., & Zheng, H. 2019. Faster Deliveries and Smarter Order Assignments for an On-Demand Meal Delivery Platform. *SSRN Electronic Journal*.

Perdikaki, O., Peng, D. X., & Heim, G. R. 2015. Impact of Customer Traffic and Service Process Outsourcing Levels on e-Retailer Operational Performance. *Production and Operations Management*, 24(11): 1794–1811.

Ramasubbu, N., & Kemerer, C. F. 2016. Technical debt and the reliability of enterprise software systems: A competing risks analysis. *Management Science*, 62(5): 1487–1510.

Rao, S., Goldsby, T. J., Griffis, S. E., & Iyengar, D. 2011. Electronic Logistics Service Quality (e-LSQ): Its Impact on the Customer ’ s Purchase Satisfaction and Retention. *Journal of Business Logistics*, 32(2): 167–179.

Schwieterman, M. A., Goldsby, T. J., & Croxton, K. L. 2018. Customer and Supplier Portfolios: Can Credit Risks be Managed Through Supply Chain Relationships? *Journal of Business Logistics*, 39(2): 123–137.

Shang, G., Pekgün, P., Ferguson, M., & Galbreth, M. 2017. How much do online consumers really value free product returns? Evidence from eBay. *Journal of Operations Management*, 53–56: 45–62.

Singh, J., Hansen, M. T., & Podolny, J. M. 2010. The world is not small for everyone: Inequity in searching for knowledge in organizations. *Management Science*, 56(9): 1415–1438.

Tereyağoğlu, N., Fader, P. S., & Veeraraghavan, S. 2017. Pricing Theater Seats: The Value of Price Commitment and Monotone Discounting. *Production and Operations Management*, 26(6): 1056–1075.

Umashankar, N., Ward, M. K., & Dahl, D. W. 2017. The benefit of becoming friends: Complaining after service failures leads customers with strong ties to increase loyalty. *Journal of Marketing*, 81(6): 79–98.

Appendix G – Institutional Review Board Protocol Approval

UNIVERSITY OF ARKANSAS

To:	Nicolo Masorgo
From:	Douglas J Adams, Chair IRB Expedited Review
Date:	04/18/2023
Action:	**Exemption Granted**
Action Date:	04/18/2023
Protocol #:	2302453760
Study Title:	the impact of last mile delivery performance on customers outcomes

The above-referenced protocol has been determined to be exempt.

If you wish to make any modifications in the approved protocol that may affect the level of risk to your participants, you must seek approval prior to implementing those changes. All modifications must provide sufficient detail to assess the impact of the change.

If you have any questions or need any assistance from the IRB, please contact the IRB Coordinator at 109 MLKG Building, 5-2208, or irb@uark.edu.

cc: David Dobrzykowski, Investigator
Brian S Fugate, Investigator

Page 1 of 1