IE UNIVERSIDAD

TESIS DOCTORAL/DOCTORAL DISSERTATION

EXAMEN DEL DESEMPEÑO DE LAS PLATAFORMAS BAJO DEMANDA A TRAVÉS DE TRES LENTES: TRABAJADOR, TIENDA Y CLIENTE

EXAMINING THE PERFORMANCE OF ON-DEMAND PLATFORMS THROUGH THREE LENSES: WORKER, STORE, AND CUSTOMER

REEJU GUHA

SEGOVIA, 2024

IE UNIVERSIDAD

TESIS DOCTORAL/DOCTORAL DISSERTATION

Examen del Desempeño de las Plataformas Bajo Demanda a través de Tres Lentes: Trabajador, Tienda y Cliente

Examining the Performance of On-Demand Platforms through Three Lenses: Worker, Store, and Customer

Reeju Guha

Doctoral Thesis Advisor: Daniel Corsten

Resumen

Las plataformas de servicios en línea han ganado terreno en la última década. Estas plataformas conectan a clientes en espera de tiempo con proveedores de servicios independientes. Entre la plétora de plataformas de servicios en línea y a la carta, el mercado de entrega de comestibles en línea, en particular, ha mostrado un rápido crecimiento y se prevé que alcance los 786.600 millones de dólares en 2024. A pesar de su creciente popularidad, las plataformas basadas en servicios en línea plantean retos operativos, como el desajuste entre la demanda y la oferta de servicios disponibles. Para aumentar la complejidad, la mayoría de las plataformas a la carta funcionan con un modelo de ”gig-contractor”, en el que los trabajadores pueden autoprogramarse, lo que aumenta la incertidumbre sobre su disponibilidad. A medida que más clientes se acostumbran a la idea de pedir comida en línea, la demanda de servicios de entrega en línea ha aumentado, al igual que la competencia entre los proveedores de servicios. Además, a menudo existe una disyuntiva entre costes y calidad a la hora de decidir si se opta por los ”almacenes oscuros” gestionados por la propia empresa para tener un mayor control sobre los tiempos de recogida y la calidad del servicio, o por recoger en las tiendas minoristas junto con otros clientes, sin coste adicional, pero a costa de unos tiempos de entrega más lentos y una calidad de servicio incierta derivada de la falta de existencias en las tiendas. Por último, está el problema de los clientes que cambian de tienda o de plataforma de alimentación en línea debido a las lucrativas promociones y ofertas y a los bajos costes de cambio. Todos estos factores convierten a las plataformas de venta de comestibles a la carta en un negocio en crecimiento, pero difícil, que plantea interesantes cuestiones de investigación. El propósito de mi investigación es examinar cómo los comportamientos inherentes a los trabajadores gig, y los factores externos específicos de la tienda de comestibles, y el consumidor final, podrían afectar el rendimiento y la supervivencia de las plataformas de comestibles en línea. En el primer capítulo, analizo los factores que afectan a la productividad y la calidad del servicio de los trabajadores autónomos que pueden autoprogramarse, y cómo esta información podría utilizarse para asignar mejor las tareas. En el segundo capítulo, examino cómo las plataformas de comestibles en línea podrían mejorar la eficiencia de la recogida de pedidos de las tiendas minoristas mediante la programación de tareas basadas en el tráfico de la tienda y la urgencia de la entrega del pedido. Por último, en el tercer capítulo, exploro el efecto del valor de la compra en línea, la calidad del servicio y la experiencia de compra en línea de los clientes sobre la probabilidad de cambio de tienda y de abandono de la plataforma (churn) de los compradores en línea.

Abstract

Online service platforms have gained traction over the past decade. These platforms connect waiting time-sensitive customers to independent service providers. Among the plethora of online and on-demand service platforms, the online grocery delivery market in particular, has shown rapid growth and is projected to reach US$786.6bn in 2024. Despite its growing popularity, online service-based platforms pose operational challenges, such as the mismatch between demand for, and supply of available services. To add to the complexity, most on-demand platforms operate on a gig-contractor model, where the workers can self-schedule, thereby increasing the uncertainty of workers' availability. As more customers are getting accustomed to the idea of ordering groceries online, the demand for online delivery services has increased, and so has the competition among the service providers. Furthermore, often there exists a cost-quality trade-off when performing decisions regarding whether to own company-managed *dark stores*' to have more control over picking times and service quality, or to pick from retail stores alongside other customers, at no additional cost, but at the expense of slower turnaround times and uncertain service quality resulting from in-store stockouts. Lastly, there is the issue of customers switching to other stores, or to other online grocery platforms owing to lucrative promotions & offers, and low switching costs. All these factors make on-demand grocery retail platforms a growing, yet challenging business, which poses interesting research questions. The purpose of my research is to examine how inherent behaviors of gig workers, and external factors specific to the grocery store, and the end-consumer, could affect the performance and survival of the online grocery platforms. In the first chapter, I discuss factors affecting the productivity and service quality of gig workers who can self-schedule, and how such information could be used to better allocate tasks. For the second chapter, I examine how online grocery platforms could improve order picking efficiency from retail stores by scheduling tasks based on store traffic, and urgency of the order delivery. Lastly, in the third chapter, I explore the effect of online purchase value, service quality, and customers' online shopping experience on the store-switching and platform-exit (churn) probability of online shoppers.

Acknowledgements

Completing a PhD thesis is a tantalizing prospect, and is not possible without the help and support of a lot of people. I would like to take this opportunity to express my sincere gratitude towards my advisor, Prof. Daniel Corsten for his invaluable advice, support, and encouragement that has guided me towards the successful completion of this document. He has been an excellent teacher and mentor, and has inspired me to strive towards excellence in teaching and research. Next, I would like to extend my sincere gratitude towards my defense committee members, Prof. Srikanth Jagabathula of the Stern School of Business, NYU, Prof. Pedro Amorim of INESC TEC, University of Porto, and Prof. Laura Zimmermann of IE Business School, IE University for sparing time out of their busy schedules to be a part of my final defense committee. I would also like to extend my gratitude to the Ph.D. program directors, Prof. Julio De Castro and Prof. Laura Maguire, along with Ms. Maria Muriel, the doctoral programs coordinator, who have ensured that all our academic needs are well taken care of. I also thank Prof. Antti Tenhiälä of IE Business School, for being a part of my Ph.D. proposal defense, and Prof. Laura Wagner of IESE Business School, for graciously accepting the offer of an external substitute for my Ph.D. final defense committee. Furthermore, I also thank all faculty members at IE Business School, whose guidance, teaching, and support has helped me appreciate research in a whole new light. I would also like to thank my fellow peers for their intellectual and emotional support during the last five years. Finally, a special thanks to my father, mother, and sister for their constant love, support, and prayers.

Contents

List of Figures

List of Tables

Introducción

Las plataformas de servicios a la carta ponen en contacto a clientes que necesitan servicios rápidos, fiables y de fácil acceso con proveedores de servicios independientes, también denominados ”trabajadores por encargo”. Estas plataformas ofrecen una gran variedad de servicios, como viajes en coche (Uber, Lyft, Bolt), reparto de comida (Dordash, Grubhub, Uber Eats), reparto de comestibles (Instacart, Deliveroo, Glovo) y muchos más. En particular, las plataformas de entrega de comestibles en línea han ganado tracción masiva en el pasado reciente. Se prevé que el mercado mundial de entrega de comestibles en línea alcance los 786.800 millones de dólares en 2024. A pesar de su creciente popularidad, las plataformas basadas en servicios en línea plantean retos operativos, como el desajuste entre la demanda y la oferta de servicios. A medida que más clientes se acostumbran a la idea de pedir productos en línea, la demanda de tales servicios ha aumentado, al igual que la competencia entre las plataformas a la carta, que se disputan una parte de la cuota en este mercado cada vez más competitivo y con escasos márgenes operativos. Para mayor complejidad, la mayoría de las plataformas a la carta operan con un modelo de ”gig-contractor”, en el que los trabajadores pueden autoprogramarse, lo que lleva a una disponibilidad incierta de los trabajadores, dificultando aún más la adecuación entre la oferta y la demanda. En estas circunstancias, es importante comprender los comportamientos y las características específicas de los trabajadores autónomos que afectan al desempeño de sus tareas.

Partiendo de esta idea, en mi primer capítulo examino cómo los comportamientos y características específicos de los trabajadores gig repercuten en la disponibilidad de trabajo y en el rendimiento de las tareas en una plataforma de comestibles en línea. Además, examino cómo estos impulsores del rendimiento podrían ser utilizados para la asignación de tareas adecuada de los trabajadores gig disponibles. La idea de este estudio tiene una motivación práctica, ya que la mayoría de las plataformas de gig-work asignan las tareas aleatoriamente al trabajador disponible más cercano. En otras plataformas, la asignación se realiza en función de las valoraciones previas o de la experiencia previa de los trabajadores. Sin embargo, estos mecanismos de asignación no suelen ser útiles, ya que los trabajadores suelen tener lagunas en la continuidad de su servicio, lo que provoca un efecto de ”olvido”, especialmente cuando los trabajadores visitan tiendas desconocidas después de una larga interrupción del servicio. Además, cuando las plataformas se expanden a nuevas zonas geográficas, es posible que los trabajadores no tengan experiencia previa (o valoraciones). Por lo tanto, investigamos si la experiencia reciente -más concretamente, la experiencia de un día- podría influir positivamente en el rendimiento de los trabajadores gig y, por lo tanto, utilizarse como parámetro de clasificación en la asignación de tareas. Desarrollamos un modelo econométrico para analizar la productividad y la calidad del servicio de los trabajadores gig, teniendo en cuenta el sesgo de selección de la muestra y la endogeneidad. A continuación, utilizamos los resultados de nuestro análisis para clasificar a los trabajadores gig en función de su experiencia previa y diaria, y desarrollamos un algoritmo de asignación de tareas para asignar el trabajo en función de la clasificación de los trabajadores gig y la complejidad de las tareas. Por último, predecimos las mejoras de rendimiento resultantes de la nueva asignación de tareas.

Mi segundo capítulo está motivado por otro problema práctico de relevancia en las plataformas a la carta. En los dos últimos años, un gran número de plataformas de entrega ultrarrápida de comestibles cerraron por completo sus negocios o se retiraron de determinados mercados, mientras que otras introdujeron importantes cambios operativos en su modelo de negocio. Sin embargo, plataformas de comestibles en línea como Instacart, Deliveroo y Glovo han prosperado. La principal diferencia entre la entrega de comestibles en línea y la ultrarrápida es que los compradores personales de las plataformas de comestibles en línea recogen los pedidos de las tiendas minoristas de comestibles y los entregan en la ubicación del cliente. Por otro lado, las plataformas ultrarrápidas (de comercio rápido) poseen sus propios microcentros de distribución, conocidos comúnmente como *tiendas oscuras*. Ambos enfoques tienen sus propias ventajas e inconvenientes. Por ejemplo, las empresas de comercio rápido tienen plazos de entrega más rápidos y se enfrentan a menos problemas relacionados con la calidad, porque las tiendas son gestionadas por la propia plataforma. Sin embargo, estas empresas incurren en costes adicionales de gestión de las ”tiendas oscuras” y de los empleados de las tiendas. Debido a los escasos márgenes operativos de este sector, incurrir en costes adicionales podría suponer una amenaza para la supervivencia del comercio rápido. Por otro lado, las plataformas de comestibles en línea realizan la recogida en las tiendas minoristas tradicionales, evitando así los costes adicionales, pero se enfrentan al riesgo de entregas más lentas y problemas de calidad del servicio, como la falta de existencias.

Por tanto, nuestra idea se centra en identificar técnicas que puedan mejorar los plazos de entrega sin incurrir en costes adicionales de gestión y mantenimiento de los ”almacenes oscuros”. En concreto, nos centramos en las ventajas de la recogida de comestibles en línea por parte de compradores personales (gig workers) de tiendas minoristas. Los ”gig workers” de las plataformas de supermercados online se encuentran con frecuencia con la presencia de clientes y otros empleados de la tienda mientras recogen pedidos de tiendas minoristas ”ocupadas”. Esto afecta a su eficiencia en la recogida debido a la interferencia de los clientes y a las largas colas en las cajas, así como a la calidad de la recogida debido a la falta de existencias y a los errores de recogida. Una alternativa viable a este problema podría ser programar la preparación de pedidos ”no urgentes” durante las horas de menor afluencia a las tiendas. Sin embargo, las horas de menor afluencia se utilizan a menudo para reabastecer las tiendas y, por lo tanto, es posible que las estanterías queden vacías y se produzcan roturas de stock. A continuación, desarrollamos un algoritmo de programación para planificar la preparación de pedidos en función de la urgencia y de las características específicas del pedido. A continuación, comparamos las mejoras previstas en términos de reducción del tiempo de preparación de pedidos y de la falta de existencias. Sin embargo, nos damos cuenta de que programar todos los pedidos con antelación no es factible debido a las características del producto (artículos refrigerados o congelados) y a la urgencia de la entrega. En tales circunstancias, proponemos dos técnicas adicionales mediante las cuales se puede reducir el tiempo de compra. En primer lugar, examinamos las ventajas de utilizar mostradores de autopago y aplicaciones de escaneado sobre la marcha para reducir el tiempo de espera en las colas y, por tanto, el tiempo de compra. A continuación, utilizamos los resultados de una prueba A/B para determinar si la adopción de un mecanismo separado de recogida y entrega puede ayudar a reducir el tiempo de entrega y los retrasos. Demostramos que nuestros mecanismos propuestos pueden ayudar a mejorar la productividad (reduciendo el tiempo de recogida y de entrega) y la calidad del servicio (reduciendo los retrasos y el número de roturas de stock).

Por último, en mi tercer capítulo, examino el comportamiento de los clientes al comprar en una plataforma de comestibles a la carta. Nuestro conjunto de datos único nos permite examinar los comportamientos de compra de los clientes durante un período de tiempo significativo. En particular, estamos interesados en examinar el efecto de los factores de ”empuje”, como el precio y la calidad del servicio, en el comportamiento de cambio de tienda de los clientes, al tiempo que incorporamos los comportamientos de los clientes y la experiencia de compra en línea en nuestros modelos. cambio de tienda es un fenómeno que se analizó originalmente en trabajos sobre comercio minorista (y marketing), en los que el coste del cambio se mide como la distancia entre tiendas, compensada con precios, surtido, etc. Nuestro trabajo difiere de los anteriores en que examinamos el cambio de tienda de los clientes en función de su tienda principal de afiliación, es decir, la tienda en la que compran la mayor parte de los artículos en un periodo de tiempo determinado. Además, en un entorno de supermercado en línea, no hay costes de cambio basados en la distancia ni fricciones temporales. En concreto, examinamos los factores que afectan a la utilidad intrínseca de los clientes hacia su tienda-marca ”principal”, y su efecto de segundo orden en el cambio de tienda principal. En concreto, nos centramos en los factores que influyen en el cliente y en la calidad del servicio en línea. Los comportamientos de compra de los clientes evolucionan a medida que adquieren experiencia de compra online. Esto afecta a la utilidad intrínseca hacia sus marcas de tiendas primarias, cuando se producen problemas de calidad de servicio como el agotamiento de existencias y las sustituciones. Nuestro objetivo, por tanto, es examinar el efecto de la experiencia en línea y de la calidad del servicio en la decisión de cambiar de tienda y de abandonar la plataforma (fuga de clientes), respectivamente.

En conclusión, nuestro trabajo examina los factores que afectan al rendimiento de las plataformas de comestibles a la carta. En concreto, examinamos cómo el trabajador autónomo, la tienda minorista y el cliente pueden afectar al rendimiento y a las futuras compras de las plataformas de comestibles en línea, y sugerimos mecanismos para mejorar el rendimiento general en términos de productividad, calidad del servicio y futuras recompras.

Introduction

On-demand service platforms connect waiting time-sensitive customers in need of rapid, easily accessible, and reliable services to independent service providers, who are also called ”gig workers”. Such platforms offer a variety of services, such as ride-hailing (Uber, Lyft, Bolt), food deliveries (Doordash, Grubhub, Uber Eats), grocery deliveries (Instacart, Deliveroo, Glovo), and many more. In particular, the online grocery delivery platforms have gained massive traction in the recent past. The worldwide online grocery delivery market is projected to reach US$786.8bn in 2024. Despite their growing popularity, online service-based platforms pose operational challenges, such as the mismatch between demand for, and supply of services. As more customers are getting accustomed to the idea of ordering products online, the demand for such services has increased and so has the competition among on-demand platforms, who are vying for a piece of the share in this increasingly competitive market with thin operating margins. To add to the complexity, most on-demand platforms operate on a gig-contractor model, where workers can self-schedule, which leads to uncertain availability of workers, thereby making demand-supply matching even more difficult. Under these circumstances, it is important to understand the specific behaviors and characteristics of gig workers that affect their task performance.

Building on this idea, in my first chapter, I examine how specific behaviors and characteristics of gig workers impact work availability and task performance in an online grocery platform. I further examine how these drivers of performance could be utilized for proper task allocation of the available gig workers. The idea of this study was practically motivated, as most gig-work platforms allocate tasks randomly to the nearest available worker. In some other platforms, allocation is done on the basis of prior ratings or, prior experience of gig workers. However, such allocation mechanisms are often not useful as gig workers frequently have gaps in the continuity of their service, thereby leading to a ‘forgetting’ effect, especially when workers visit unfamiliar stores after a long service gap. Furthermore, when platforms expand to new geographies, gig workers might not have prior experience (or ratings). We therefore investigate whether recent experience - more specifically, within-day experience, could positively impact the performance of gig workers, and therefore be utilized as a ranking parameter in task allocations. We develop an econometric model to analyze gig workers’ productivity and service quality, while accounting for sample selection bias and endogeneity. We then utilize the results from our analysis to rank gig workers based on prior and within-day experience, and develop a task allocation algorithm to allocate work based on gig workers’ rank, and the complexity of the tasks. Finally, we predict performance improvements resulting from the new task allocations.

My second chapter is motivated from another practical problem of relevance in on-demand platforms. The last couple of years saw a great number of ultrafast grocery delivery platforms either completely shut their businesses down or exit from certain markets, while others made significant operational changes in their business model. However, online grocery platforms such as Instacart, Deliveroo, and Glovo have flourished. The major difference between online, and ultrafast grocery delivery is that personal shoppers of online grocery platforms pick orders from grocery retail stores and deliver it to the customer’s location. On the other hand, ultrafast (quick commerce) platforms own their own micro fulfillment centers, commonly known as *dark stores*. Both approaches have their own advantages and disadvantages. For example, quick commerce companies have faster turnaround times and face fewer quality-related issues, because the stores are managed by the platform itself. However, these companies incur additional costs in managing the ‘dark stores’ and the in-store employees. Owing to the thin operating margins in this industry, incurring additional costs could pose a threat to the survival of quick commerce. On the other hand, online grocery platforms perform picking from traditional retail stores, thereby avoiding the additional costs, but face the risk of slower deliveries and service quality issues such as stockouts.

Central to our idea, therefore, is to identify techniques that could improve turnaround times without incurring additional costs of managing and maintaining ‘dark stores’. Specifically, we focus on the benefits of online grocery picking by personal shoppers (gig workers) from retail stores. Gig workers of online grocery platforms frequently encounter the presence of customers and other in-store employees while picking orders from ‘busy’ retail stores. This affects their picking efficiency due to customer interference, and longer checkout queues; and the quality of picking due to out-of-stocks and picking errors. A viable alternative to this problem could be to schedule the picking of ‘non-urgent’ orders during hours that are characterised by low store traffic. However, ‘less-busy’ store hours are often utilized for store replenishment, and therefore, one might encounter empty shelves and out-of-stocks (OOS). We therefore match similar order baskets picked during busy vs less-busy periods to establish the value of picking during less-busy periods. Next, we develop a scheduling algorithm to schedule order picking based on urgency and order-specific characteristics. We then compare the predicted improvements in terms of reduction in order picking time and out-of-stocks. However, we do realize that scheduling all orders ahead of time is not feasible due to product characteristics (chilled or frozen items), and delivery urgency. In such circumstances, we propose two additional techniques through which shopping time can be reduced. First, we examine the benefits of using self-checkout counters and *scan-on-the-go* apps to reduce queuing time, thereby reducing the shopping time. Next, we utilize the results from an A/B test to determine whether adopting a separate picker and deliverer mechanism can help reduce delivery time and delays. We demonstrate that our proposed mechanisms can help improve productivity (by reducing picking time and delivery time), and service quality (by reducing delays, and number of stockouts).

Finally, in my third chapter, I examine customers' behaviors while shopping from an on-demand grocery platform. Our unique dataset allows us to examine customers' purchasing behaviors over a significant period of time. In particular, we are interested in examining the effect of 'push' factors such as price, and service quality on customers' store-switching behavior, while incorporating customers' behaviors and online shopping experience in our models. *Store switching* is a phenomenon that was originally examined in retailing (and marketing) papers, where the switching cost is measured as distance between stores, traded-off with prices, assortment etc. Our work differs from prior work in that we examine customers' store-switching in terms of their primary store of affiliation - i.e., the store from which they purchase the largest share of items in a given time period. Moreover, in an online grocery setting, there are no distance-based switching costs or temporal frictions. Specifically, we examine the factors affecting customers' intrinsic utility towards their 'primary' online store-brand, and its second order effect on primary store switching. In particular, we focus on customer-specific drivers, and online service quality drivers. Customers' shopping behaviors evolve as they gain online shopping experience. This affects the intrinsic utility towards their primary store brands, when service quality issues such as stockouts and substitutions occur. Our objective, therefore, is to examine the effect of online experience and service quality on the decision to switch stores, and to exit the platform (customer churn), respectively.

In conclusion, our work examines factors affecting the performance of on-demand grocery platforms. Specifically, we examine how the gig worker, the retail store, and the customer can affect performance of, and future purchases from online grocery platforms, and suggest mechanisms to improve the overall performance in terms of productivity, service quality, and future repurchase.

Chapter 1

The Role of Within-Day Learning on Gig Workers' Performance and Task Allocation

1.1 Abstract

Online platforms operating on a gig-work model rely on self-scheduling gig contractors to meet real-time demand, while maintaining the desired productivity and service quality levels. Such platforms either adopt a *broadcast* mechanism, where workers have the autonomy to select orders, or a *dispatch* mechanism, where the platform assigns orders to workers. In most platforms, however, tasks are randomly allocated based on the proximity of the worker to the store, or on the basis of prior ratings of workers. However, such allocation mechanisms are often not useful as gig workers frequently have gaps in the continuity of their service, thereby leading to a 'forgetting' effect, especially when workers visit unfamiliar stores after a long service gap. Furthermore, when platforms expand to new geographies, gig workers might not have prior experience (or ratings). Drawing from theories of *learning*, and *flow*, we investigate whether *within-day experience* can positively impact workers' performance, and therefore be utilized as a ranking parameter in task allocations. Utilizing data from an online grocery platform, we develop an econometric model to analyze gig workers' productivity and service quality based on their within-day experience, while accounting for sample selection bias and endogeneity. Our findings reveal that as within-day experience increases, both productivity and service quality improves. Since we aim to allocate work based on task characteristics, we examine the effect of order batching and task complexity on performance. We observe that when workers batch orders, within-day experience reduces delays, but also reduces picking productivity and item substitutions. We further observe, for complex tasks, higher same-day experience is beneficial up to a threshold, beyond which it improves picking productivity but reduces substitutions. Utilizing results from our analysis, we rank gig workers based on prior and same-day experience and develop a task allocation algorithm to allocate tasks based on gig workers' rank, and the complexity of the tasks. We calculate predicted improvements in productivity and service quality from the new task allocations. The predictions demonstrate that allocating gig workers to tasks by accounting for their behaviors and skills, along with the complexity of the tasks, while limiting order batching leads to performance improvements.

1.2 Introduction

Online service platforms have gained traction over the past decade. These platforms connect waiting time-sensitive customers to independent service providers (Taylor, 2018; Benjaafar et al., 2022; Benjaafar and Hu, 2020; Guha, 2023). Such platforms offer a variety of services, such as ride-hailing (Uber,Lyft, Bolt), food deliveries (Doordash, Grubhub, Uber Eats), grocery deliveries (Instacart,Deliveroo, Glovo), and many more. According to Statista (2024), the worldwide online grocery delivery market alone, is projected to reach US$786.6bn in 2024. In recent years, on-demand grocers started offering ultra-fast deliveries which promise 10-15 minute deliveries (Cheng, 2021a), thereby improving the online grocery shopping experience even further. Despite its growing popularity, gig work platforms pose operational challenges, such as the mismatch between demand for, and supply of available services (Benjaafar and Hu, 2020; Allon et al., 2023; Xu et al., 2023; Guha, 2023). Since gig contractors are not directly under the platform's control and can self-schedule, having rapid turnaround times is often not feasible, as it requires workers to be available at the right time to serve consumers at short notice (Guda and Subramanian, 2019; Guha, 2023). Moreover, even with available workers, platforms often struggle with delivery task allocation, which in turn affects on-time task completion (Sharma, 2020). It is therefore crucial for managers to understand the worker behaviors that could impact order-level productivity and service quality under this business model. This paper seeks to investigate how learning by doing can help gig workers improve their task performance, and how managers could utilize this information to better allocate tasks.

Recent work in the sharing (gig) economy setting has examined gig workers' labor decisions such as when and for how long do such workers make themselves available for work (Allon et al., 2023), and how incentives and penalties affect these working decisions (Xu et al., 2023). An aspect that has been ignored in this literature is how gig worker behaviors might impact individual task productivity and service quality. Following this, another relevant aspect for practitioners is how to allocate tasks to available workers in a way that improves overall task performance. Most on-demand grocery platforms either adopt a broadcast mechanism, where workers have the autonomy to select orders, or a dispatch mechanism, where the platform assigns orders to workers (Dai et al., 2022). However, when it comes to task suggestions (or assignments), the objective of platforms is to deliver orders within the promised time, while minimizing delivery cost (Glovo, 2021; Paul, 2022). In doing so, the algorithms do not take workers' behaviors and skills into account, which could help provide a better quality of service. An exception to this approach is order allocation based on the overall ratings of personal shoppers, adopted by Instacart (Helling, 2023). However, such an approach can be problematic as new workers might not have sufficient ratings to demonstrate their abilities. Moreover, since customers can select not to rate at all, a higher (lower) rating might not be a true representation of the workers' skills. Another often suggested approach in the literature is the task experience (or familiarity)-based allocation of individuals to tasks. For example, Espinosa et al. (2007) posit that task familiarity is more beneficial with more complex tasks. However, platforms frequently expand to new geographies and settings where the available gig workers might not have prior experience. Moreover, simply allocating tasks to experienced workers could cause an experience disparity, as new workers would not gain sufficient delivery experience. Another concern raised by Dai et al. (2022) is that in a gig-work setting, individual experience might have adverse effects on worker performance, especially at the exploration stage. Thus, allocating (or suggesting) tasks to workers based on experience alone, is unlikely to yield significant performance benefits. We therefore investigate whether recent experience - more specifically, within-day experience, could positively impact the performance of gig workers, and therefore be utilized as a ranking parameter in task allocations.

To examine how managers of sharing platforms can effectively suggest (or assign) tasks to gig workers, it is important to consider how gig workers make labor decisions (Allon et al., 2023). Such decisions capture the intrinsic and extrinsic motivations of workers (Xu et al., 2023), which in turn affects workers' task performance. As workers perform additional tasks, they benefit from the phenomenon of *learning by doing*. Individual learning can take place both within the same day, or across days, as the worker gathers experience. There is an abundance of literature on the role of individual learning (and experience) in affecting workers' task performance. See, for example, (Reagans et al., 2005; Huckman and Pisano, 2006; Narayanan et al., 2009; Kc and Staats, 2012; Ramdas et al., 2018; Batt and Gallino, 2019; Bavafa and Jónasson, 2021). However, only a handful of studies focus on within-day learning on individual task performance. Among existing studies Quinn (2005) explain that within the day, as workers' indulgence increases, their sense of control over an activity, and the confidence in their ability to perform the said activity increases, resulting in improved performance (Guha, 2023). In another study, Staats and Gino (2012) find that same-day, same-task experience has a positive effect on worker productivity and could lead to a reduction in task completion time. None of the studies, however, examine the effects of within-day learning in a gig-work setting, where although some work is repetitive, each task differs from the other in terms of complexity, time, and effort required. Directly applying findings from prior studies with a traditional workforce setting to gig platforms could lead to misleading conclusions (Agrawal et al., 2015; Guha, 2023) because gig workers have the flexibility to select their work hours and which tasks to work on, with limited supervision from managers (Dai et al., 2022). Furthermore, since gig workers are likely to select hours that provide them sufficient rest and breaks for personal chores, their performance in individual tasks is less likely to be hampered due to loss of concentration, demotivation, or fatigue (Kelliher and Anderson, 2010; Cousins and Tang, 2016; Guha, 2023). As a result, previous findings which posit the possibility of an inverted-U relation between same-day learning and productivity might not hold true in this new setting. Therefore, the first goal of this work is to investigate the effects of same-day experience on gig workers' productivity and service quality in individual tasks. We measure productivity using picking time, and service quality with order delays, and item substitution for stocked-out items when customers requested a substitution.

Recent studies demonstrate that longer voluntary work reflects gig workers' motivation to work, owing to specific behaviors such as an earnings target-setting behavior (Sheldon, 2016; Chen et al., 2019; Xu et al., 2023; Guha, 2023), which positively affects the productivity rate, and intrinsic motivation, which positively affects the service quality (Xu et al., 2023; Dai et al., 2022). However, it is quite likely that workers might become less attentive to details as their working time increases. Thus, inferring from the results documented in the prior studies, workers' within-day experience may have both positive and adverse effects on performance. To the best of our knowledge, Dai et al. (2022) is the only study that has examined how prior experience in a gig-work setting, affects service quality. Our work differs from this study as we examine the effect of gig workers' within-day experience on individual task performance, as opposed to the effect of prior experience on aggregate task performance across all orders delivered in a day.

The second goal of this work is to understand how to utilize gig workers' within-day learning and prior task experience to allocate tasks. Our objective is to examine whether within-day experience could be useful in handling specific tasks that are more demanding. We first examine *order batching*, which is a common practice in on-demand food and grocery delivery platforms, where the order picker picks and delivers multiple orders together (Guha, 2023). Gig workers' order batching decisions may have a direct impact on task productivity and service quality. For example, order batching from the same store could reduce the picking time per order, as it could reduce both travel and search costs (Batt and Gallino, 2019). It could also lead to improvements in service quality by reducing delays. On the other hand, batching could reduce productivity by increasing the average picking time per order, especially if the orders are accepted at different times within the same hourly slot. Similarly, the income-targeting behavior of gig workers (Allon et al., 2023) can make them more productive at the expense of service quality. It is therefore crucial to understand whether higher same-day experience is beneficial when orders are batched.

Next, we examine the effect of task complexity and gig workers' discretion in complex tasks on performance. In our setting, we measure task complexity based on the number of stockout items for which a substitution was requested. Stock-out substitutions are not only time-consuming, but also involve higher discretion and skill. For example, in our setting, if a worker encounters a stockout, she needs to first check whether the customer has requested a substitution or not. If yes, the worker needs to either call the customer and suggest alternatives or use their discretion to select a suitable substitute. Prior research has revealed that as task complexity increases, it is better to have greater experience at the individual worker level (Huckman and Staats, 2011). However, whether it is better to allocate a complex task to a gig worker who might have the same level of overall experience, but higher (lower) within-day experience, remains unexplored. Understanding the effect of task complexity and gig workers' discretion could be crucial in allocating tasks, especially when workers do not have sufficient prior experience or, all available workers have a similar level of experience.

Our final objective is to devise a task allocation mechanism that increases gig workers' performance. We propose a ranking methodology based on gig workers' within-day experience and prior task experience, to rank available gig workers in each hourly slot. We then develop an allocation algorithm to assign complex tasks to higher-ranked workers, while accounting for order batching possibilities. We incorporate multiple constraints to replicate the realities of our setting.

We test our research questions using a comprehensive dataset from an on-demand grocery platform. The gig-work setting has unique empirical challenges as compared to traditional settings, for example, gig workers can choose which slots they want to work in, leading to potential self-selection bias in the estimation. To tackle this challenge, we employ a two-stage Heckman selection model to model workers' slot selection decisions in the first-stage *choice equation* and examine the effects of work availability and working time on performance in the second-stage *level equation*.

This study makes multiple contributions to theory and practice. First, we establish a causal relationship between gig workers' same-day experience and task performance. We find that as within-day experience increases, both order-level productivity and service quality improves. Second, by investigating the moderating effects of order batching, task complexity, and worker discretion in complex tasks, we provide insights into when the benefits of longer same-day work on task performance are most pronounced. We find that longer same-day work hours bring more prominent productivity benefits when order batching is present at low levels, task complexity is moderate, and worker discretion in complex tasks is also moderate. Finally, predictions from the new task allocation algorithm, based on our econometric models, demonstrate improvements in gig worker productivity and service quality.

1.3 Theory and Related Literature

Our work is related to three streams of literature: (i) gig worker behavior in sharing platforms, (ii) the impact of working time flexibility, and within-day learning on workers' task performance, (iii) the effect of task characteristics, such as, task batching and task complexity on performance.

1.3.1 Gig worker behaviors in sharing platforms

Our work is closely related to studies on gig worker behaviors in sharing platforms. Past studies have examined various aspects of gig worker behaviors and their impact on gig worker availability and performance (Chen and Horton, 2016; Sheldon, 2016; Chen et al., 2019; Hall et al., 2021; He et al., 2021; Thakral and Tô, 2021; Allon et al., 2023; Xu et al., 2023; Dai et al., 2022). Primarily, the focus of the investigations has been examining the impact of earnings on gig worker behaviors. In particular, Sheldon (2016) investigates the effect of wages on workers' decision to work and their working time. Thakral and Tô (2021) and Chen et al. (2019) examine workers' income targeting behavior, and achieving the desired targets, on the decision to work. Allon et al. (2023) study the effect of behavioral and economic factors on gig workers' decisions to work, and the duration of work. Xu et al. (2023) examine how incentive schemes of sharing platforms, such as earnings, ratings, and penalties, affect the working decisions of gig workers. However, so far none of the existing studies have examined whether frequent gaps in service of gig workers could cause a 'forgetting' effect, and therefore accentuate the value of recent (or within-day) experience. Specifically, none of the studies examine how prior experience and within-day experience could work in-tandem to improve gig workers' productivity and service quality in individual tasks. To the best of our knowledge, Dai et al. (2022) is the only study that examines how prior experience affects gig workers' performance at an aggregate level. In this study, we focus on individual tasks and within-day experience. By focusing on individual tasks, we are better able to account for task characteristics (e.g., picking and delivery time of individual tasks, batching, complexity, and other order-specific information) that could affect the productivity and service quality of orders delivered.

1.3.2 Same-day learning, Working time flexibility, and Performance

Our primary objective is to examine the effect of within-day (learning) experience on gig workers' performance in individual tasks. Influential literature has examined the effect of within-day work through different lenses. Most notable are the studies on the effect of (same-day) working time on performance. Past studies that examine the effect of working time on performance, reveal contradictory findings. Feldstein (1967) finds that longer hours improve productivity if a worker faces fixed set-up costs, or if longer hours lead to a better utilization of capital goods. However, recent work by Pencavel (2015) reveals that the marginal effect on productivity per worker, with each additional hour of work, starts decreasing due to worker fatigue, which is likely to set in after a specific duration of hours worked (Guha, 2023). We refer to Kc (2020) for a detailed review of studies demonstrating the detrimental effects of fatigue and overwork on performance. None of these studies consider the positive effects of within-day experience, which takes place with an increase in same-day learning on worker performance. Among another branch of studies, Quinn (2005) demonstrates that within-day experience not only provides mental benefits, but also physical benefits. The author explains that as individuals start executing tasks, in addition to gaining control and confidence over the tasks performed, they gain muscle memory, which helps improve productivity. Along similar lines, Staats and Gino (2012) hypothesize that same-day, same-task experience has a positive effect on worker productivity. Among other studies, Batt and Gallino (2019) demonstrate that picker experience could reduce search time for items, leading to improved productivity. Although the authors use cumulative experience instead of recent (within-day) experience, such a finding could also be expected as workers gain higher same-day similar-task experience (Guha, 2023). In addition to the effect on *average* performance, prior experience (or same-day experience) can affect performance *variance* as well. Bavafa and Jónasson (2021) and Dai et al. (2022) find that experience can reduce variance in task completion time and delay time, respectively. To reconcile the contradictory views on the impact of same-day experience on performance, we examine whether working time flexibility of workers can help reduce the detrimental effects of fatigue, resulting from longer same-day work.

Since all workers in our setting are gig workers who have full flexibility to select their work hours, such flexible work hours can directly or indirectly impact within-day task performance. For example, Kelliher and Anderson (2010) and Golden (2012) reveal that providing employees with schedule flexibility affects productivity positively, because employees make a greater effort in exchange for working in an environment that caters to their preferences. Golden (2012) posits that greater control over the timing of work helps workers alleviate the detrimental effects of long hours such as fatigue and demotivation and leads to improved performance. Among studies examining the effect of working time on service quality, Dai et al. (2015) found that healthcare workers become less compliant with handwashing rules over the duration of their shift, indicating that a worker's attention towards service quality might shift as she works longer hours. Among other studies, Lu and Lu (2017) find that the introduction of overtime laws reduced the quality of care provided by nurses. This reduction in quality was caused by changes in staffing policies of permanent and contractual nurses (Guha, 2023). Contrary to the previous studies that demonstrate a detrimental effect of working time on quality, Kesavan et al. (2022) find that increasing workers' control over their schedule provides workers with a greater say in their work hours. They further demonstrate that increasing the average weekly work hours advances “distributive justice” for part-time workers by offering workers a greater share of available hours and pay, which in turn leads to improved service quality and performance (Guha, 2023). Drawing from a recent study by Xu et al. (2023), we expect that some workers are more extrinsically motivated, while others are intrinsically motivated. Gig workers with higher extrinsic motivation are more likely to remain productive, to increase their earnings, as same-day work duration increases. On the other hand, gig workers with higher intrinsic motivation are more likely to focus on quality, as higher service quality is likely to get rewarded with better ratings. Therefore, we expect that as the within-day experience of gig workers increases, it leads to an improvement in picking productivity, measured as reduced pick times; and service quality, measured as reduced order delays, and increased item substitutions for stockouts.

Hypothesis 1: *As within-day experience increases, gig workers’ productivity and service quality in individual tasks increases.*

1.3.3 Order batching and task performance

One of the objectives of our research is to allocate workers with higher same-day and prior experience to more challenging tasks. Therefore, it is crucial to understand the effect of within-day experience on performance when order batching is present. Past studies that examined the implications of multitasking on the productivity of individual workers (Kc, 2020) have found evidence of both positive and negative impacts of multitasking on performance (Guha, 2023). For example, Kc (2014) found evidence of the detrimental impact of multitasking on physician productivity and the task quality. However, the study reveals that a small increase in multitasking could enhance productivity and service quality, by reducing idle time and increasing worker engagement (Guha, 2023). In another study, Ibanez et al. (2018) examine the task batching behavior of radiologists and found that prioritization and batching of tasks led to a detrimental effect on both productivity and service quality.

However, none of these studies have examined the effect of task batching in a gig-work setting, and whether it is beneficial when workers' within-day experience increases. To understand this effect, we first consider the direct effect of order batching on picking and delivery performance. Although batching could decrease the average picking time per order by reducing the travel and search costs (Batt and Gallino, 2019), it could also increase the average picking time if the orders arrive at different times within the same hourly slot, or have distinct items. The detrimental effect of batching would be more pronounced if the number of batched orders is higher, as any initial improvements in picking time and delays would fade away as the number of batched orders increases (Kc, 2014; Lupien et al., 2007; Guha, 2023). This, in turn, could increase delays by prolonging both picking and delivery times. Furthermore, as the number of batched orders increases, the chances of incorrectly substituted items could increase, leading to reduced service quality. Therefore, we might expect an initial boost in task performance with an increase in order batching, but the positive effects might fade away as the number of batched orders increases. We now examine the effect of longer same-day experience when batching is present.

As workers' within-day experience increases, they gain more control over the tasks, which helps them adapt their picking technique in batched orders, thereby improving productivity and reducing delays. Dai et al. (2022) demonstrate that experience gained through batched orders can help improve both productivity and service quality. Although the findings of this study are not directly applicable to our work because gig workers in our setting are assigned orders by the platform and do not have full autonomy in order selection, we expect that within-day experience gained through batched orders would reduce delays. Moreover, once workers gain sufficient same-day experience, they exploit their learning to better estimate order processing times, which in turn leads to improved productivity (Guha, 2023). In terms of service quality, we expect intrinsically motivated workers to be attentive toward item substitutions. However, extrinsically motivated workers could be less attentive to details, thus reducing service quality. In some platform settings, workers are penalized for receiving poor ratings (Xu et al., 2023). However, in this case, gig workers are not penalized for poor performance. As a result, the service quality is more likely to suffer as workers with higher same-day experience deliver batched orders. Therefore, we conjecture a different effect on each of the three measures of worker performance:

Hypothesis 2: *Higher within-day experience reduces stockout substitutions, when orders are batched.*

Hypothesis 2b: *Higher within-day experience improves picking productivity and reduces delays, when orders are batched.*

1.3.4 Task complexity, discretion, and performance

Next, we focus on the effect of same-day experience on performance when task complexity is high. Remember that task complexity in our setting is defined as the number of stockout items in an order for which a substitution was requested. The higher the count of items to be substituted, the more complex the task, as order pickers would need to either call the customer and suggest available substitutes or use their discretion to perform substitutions. In either case, it involves additional time and thus has a direct impact on picking time (productivity). Moreover, substitution is an important quality measure, as customers could easily switch to a different platform if they consistently encounter stockouts or improper substitutions in the orders received. Since gig workers do not receive formal training on the job, learning from same-day and prior experience helps becomes crucial while handling complex tasks involving substitutions. Therefore, we investigate whether same-day experience can help workers better manage such tasks.

Existing research on the impact of task complexity on productivity reveals mixed findings. Huckman and Staats (2011) reveal that when complexity increases, greater experience at the individual-level is more beneficial than that at the team-level. Kc et al.

(2020) investigate physicians' discretion while performing complex tasks and find that although task completion preference of physicians improves the throughput volume within a shift, when the throughput volume is adjusted for complexity, selecting less complex tasks leads to worse throughput volume performance (Guha, 2023). To understand the effect of within-day experience on performance when task complexity is higher, we need to first understand the mechanisms of how gig workers' motivation affects the productivity and quality of service delivered. Xu et al. (2023) and (Allon et al., 2023) demonstrate that gig workers' motivation to work longer is driven by both intrinsic motivation, which results from past ratings and penalties; and extrinsic motivation, which results from actual or possible earnings on a given shift or day (Guha, 2023). It could therefore be inferred that gig workers who have higher within-day experience tend to have higher extrinsic and (possibly) higher intrinsic motivation, which is likely to improve their productivity and quality of service. Drawing from these studies, we could infer that workers with higher within-day experience will perform order picking faster than their counterparts, when the task is more complex (Guha, 2023). However, it is possible that as same-day experience increases beyond a threshold, the effect of fatigue will become more prominent (Golden, 2012; Kc, 2020). As a result, the motivation of workers to be attentive toward the service quality might decrease, negatively affecting item substitutions.

Next, we discuss the discretionary behavior of gig workers when faced with complex tasks. Ibanez et al. (2018) study the discretionary behavior of radiologists and find that when doctors deviate from the prescribed sequence of tasks, it seldom yields productivity gains. Specifically, when doctors use their discretion to prioritize less complex tasks, it could often lead to losses in productivity. Past studies in settings involving gig contractors have also investigated the discretionary behavior of gig workers in selecting tasks (Allon et al., 2023; Xu et al., 2023; Dai et al., 2022). However, none of these studies investigate the impact of gig workers' discretionary behavior on task performance, especially when the tasks are complex in nature. In our setting, when the customer requests that an item substitution be made without calling, it requires the worker to use her discretion to substitute. We expect that while performing complex tasks involving discretion, the longer same-day experience of gig workers could have opposing effects on productivity, and service quality (Guha, 2023). We conjecture that when gig workers with longer same-day experience use their discretion, they are quicker in making substitutions as it does not require them to call the customer, thereby saving time and improving productivity. On the other hand, the picker is more likely to make inaccurate or no substitutions at all, because they might be unsure of what the customer wants, or they might be driven by their extrinsic motivation to complete the task faster (Guha, 2023). Therefore:

Hypothesis 3: *Same-day experience increases the picking productivity of gig workers while delivering complex tasks, but reduces the quality of service delivered.*

Hypothesis 4: *Same-day experience increases the picking productivity of gig workers while delivering complex tasks using their discretion, but the reduces the quality of service delivered.*

Note that since task complexity affects order picking time (and stockout item substitutions), we do not interpret its effect on order delays, as we already control for picking time in the delay equation. Table 1.1 presents a summary of the hypotheses below:

Table 1.1: Summary of Hypotheses: Expected Directions

	(1) pickingtime	(2) delay	(3) substituted_when_reqd
H1 (OSF)	$\downarrow$	$\downarrow$	$\uparrow$
H2 (OSF*batching)	$\downarrow$	$\downarrow$	$\downarrow$
H3 (OSF*complexity)	$\downarrow$	-	$\downarrow$
H4 (OSF*discretion)	$\downarrow$	-	$\downarrow$

$^{\uparrow}$ *increases*, $^{\downarrow}$ *decreases*

1.4 Research Setting and Data

We test our hypotheses using data collected from a leading online grocery delivery platform in Europe. Please refer to Appendix, Figure A.1 for a detailed description of the business model of on-demand grocery platforms. Our data set includes roughly 1.68mn orders delivered by 6,303 gig workers from 1,782 stores (98 unique store brands) across Italy from 1^{st} October 2020 to 30^{th} September 2021 for a period of 12 months. Appendix, Figure A.2 provides a brief overview of the landing page of a personal shopper. In this section, we first present the details of our data set and then introduce the key variables used in our estimation.

1.4.1 Data Overview

Our consolidated data set has been prepared from five different data parts – (i) Order data, which includes information regarding unique orders, customers, order pickers, stores, city, timestamps (first item pick time, last item pick time, paid at counter time, final delivery time, and delivery deadline time), compensation, call duration with customer, order batching information, and Net Promoter Score (NPS); (ii) Order item data, which includes category name and brand information regarding each item in the order, information on whether the item requested was found (purchased the requested product), replaced (substituted the product with an alternative), or missing (stocked-out), item price, quantity of each item requested and found, and whether the customer had requested a substitution with or without call for each individual item, in case the item was stocked out; (iii) Order picker data, which includes picker, all orders delivered by the picker along with the picking date, gender, age, and country of birth; (iv) Customer data, which includes customer id, all orders placed by the customer along with the date, customer postal code, gender, and age; (v) Store data, which includes store id, store brand id, store postal code, store size, and whether the store is a partner. Merging the five parts, we create a consolidated data set.

Our data set contains orders picked and delivered by gig workers using our focal platform during the estimation period. We further remark that the workers did not take any orders from other stores on the platform, except for these 98 store brands. Since our focal platform is one of the largest on-demand grocery delivery platforms in Italy, and these 98 store brands cover almost all major store brands in Italy, we would expect only a small proportion of orders to be taken from other stores using other platforms. Although it is unlikely that workers gain experience by shopping from other store brands not listed by our focal platform, workers may gain unobserved additional experience by shopping from the same store brands for other platforms. To further alleviate such concerns, we use instrumental variables approach, illustrated in section 4.2.

Our original data consists of 2,913,500 orders delivered from 1st January 2019 to 4th October 2021. Due to the lack of gig worker earnings data in the period of 1st Jan 2019 to 1st October 2020, we exclude all observations in this period. Moreover, the data in this period is prone to errors due to several A/B tests such as picking and delivery being performed by two separate workers, the advent of Covid-19 during the March to June 2020 period, etc. Of the remaining 1,695,248 observations, 12,611 orders delivered outside Italy in countries such as the Czech Republic, Poland, and France, with no information regarding the cities of operation were excluded. Finally, we exclude 4,113 orders for which the pick start or pick end timestamp was missing. The final data set consists of 1,678,524 order observations. To account for workers' self-selection of hourly slots, following Allon et al. (2023)'s approach of drivers' shift selection, we expand the order-level data into a courier-day-slot panel, where *courier_id* is the unique worker id. Each day is split into 14 hourly slots between 7 a.m. till 9 p.m. during which gig workers pick and deliver orders. Workers can select the slots in which they prefer to work and are allocated orders placed in those slots. The worker-day-slot panel has a total of 6,046,793 observations to investigate workers' daily slot-wise working decisions. Finally, we exclude the first day's observations because we need to include lag variables in our estimation; see Section 4.1 for details.

1.4.2 Key Variables

In this subsection, we discuss the key variables in the model. We begin with the dependent variables. Table 1.2 reports the summary statistics of the key variables used in the 2nd stage level equation.

Table 1.2: Descriptive Statistics: Level Equation

Variable	Description	Mean	SD	Min	Max
pickingtime	Picking time (in minutes) of order t	24.72	15.12	1.00	83.01
delay	Delay (in minutes) of order t	6.51	17.04	0.00	111.41
substituted_when_reqd	Number of stockout items substituted when requested	1.84	2.24	0.00	48.00
OSF	Same-day orders delivered so far before order t	1.97	2.22	0.00	25.00
OSF_other	Orders by other workers in same city-date-slot as worker i	2.07	1.76	0.00	14.00
experience	log of worker i's prior task experience before the focal day, d	5.48	1.47	0.00	8.08
advancetime	Time between pick start and delivery deadline	86.29	38.37	0.00	174.76
num_item	Number of items in order	28.28	14.91	1.00	100.00
num_substitution	Number of substitutions in order t	1.84	2.24	0.00	48.00
call_duration	Call duration (in minutes) during order t	0.05	0.37	0.00	11.83
storefamiliarity	log of worker i's experience in the same store as order t	3.93	1.52	0.69	7.90
store_size	Size of the store (in '000 sqm)	2.24	2.01	0.11	14.50
num_batched	Number of batched orders	0.40	0.83	0.00	10.00
ccomplexity	Stockout items for which substitution was requested	2.79	2.89	0.00	14.00
cdiscretion	Items for which substitution was requested without call	0.85	1.75	0.00	9.00

Dependent Variables. Our estimation is based on the Heckman selection model, which consists of two steps: the choice equation (first stage) and the level equation (second stage); see Section 4.1 for more details. For the first stage choice equation, the dependent variable is a dummy variable, $worked_{i,s,d}$ which equals 1 if worker *i* works on slot *s* of day *d* (Guha, 2023). For the second stage level equation, we consider three key dependent variables related to workers' task performance:

Picking time: Order picking time, a measure of worker productivity, is measured as the period (in minutes) between the first and the last item of the order, and denoted as $pickingtime_{i,t}$, where *t* denotes the focal task. For batched orders, the picking times overlap between orders. We, therefore, calculate the picking time of a batched order as the difference between the pick-end time of the last batched order, and the pick-start time of the first batched order, multiplied by the ratio of the number of items in the focal order and the total number of items in all orders combined. Specifically, let us assume there are 2 batched orders (A and B). Order A contains 10 items, and Order B contains 15 items. The pick-start time of order A is 9:01 a.m., and the pick-start time of order B is 9:20 a.m., while the pick-end time of order A is 9:34 a.m. and the pick-end time of order B is 9:36 a.m. So the picking time of order A will be (9:36 a.m. - 9:01 a.m.)*(10/(10+15)), i.e., 35*(10/25), or 14 minutes (instead of 33 minutes), and for order B it will be 21 minutes.

Delay: Delay time per order (in minutes), a measure of service quality, is calculated as the additional time taken after the delivery deadline, and denoted as $delay_{i,t}$. Delay equals 0 if the order was delivered before the deadline.

Substituted when requested: Stockout items substituted when requested by the customer, another measure of service quality, is measured as the count of items substituted when an item was stocked out and a substitution was requested, and is denoted as $substituted\_when\_req_{i,t}$

Independent Variables. In this section, we discuss our independent variables – within-day experience, order batching, complexity, discretion, and other control variables.

Within-day experience: The key independent variable in our estimation is the experience of a worker *i* before task *t* performed in slot *s* on the day *d*. Following (Bavafa and Jónasson, 2021; Staats and Gino, 2012), we utilize the cumulative number of orders delivered by worker *i* on the day *d* before task *t* as the primary measure of same-day experience, denoted by the variable, orders so far ($OSF_{i,t}$). Later in our robustness check, we also use the cumulative number of hours worked on the same day before the focal task ($HSF_{i,t}$) to rerun our main estimation.

Order Batching: The variable *num_batched* is the number of orders batched together. An order is said to be batched if multiple orders are picked and delivered simultaneously by the gig worker *i* from the same store in the same slot, or two consecutive slots of a given day. Recall that since an order can be picked in one slot and delivered in the next, it is possible that the picking of two or more orders *started* simultaneously or after some time in a specific slot and *ended* in the next slot.

Complexity: We define an order to be more complex if the number of stocked-out items in an order is higher. In our main analysis, we use $ccomplexity$, which is a (top 1%) winsorized measure of $numstockout\_reqsub$. We winsorize the variable as extreme values could influence the impact of the variable. In the online appendix tables A.4, A.5, A.6, we measure complexity as a categorical variable that takes the value of high, medium, or low. Complexity is *low* if $numstockout\_reqsub \leq 2$, which corresponds to the 50th %ile; *medium* if $numstockout\_reqsub > 2$ and ≤ 7, which corresponds to the 90^{th} %ile; and *high* if $numstockout\_reqsub > 7$.

Discretion: An order requires more discretion from the gig worker when an order has a higher number of stocked-out items for which a substitution was requested without calling the customer. In our main analysis, we use $cdiscretion$, which is a (top 1%) winsorized measure of $stockout\_sub\_withoutcall$. In Online Appendix tables A.4, A.5, A.6, we measure discretion as a categorical variable that takes the value of high, medium, or low. Discretion is *low* if $stockout\_sub\_withoutcall \leq 1$, which corresponds to the 75th %ile; *medium* if $stockout\_sub\_withoutcall > 1$ and ≤ 5, which corresponds to the 95th %ile; and *high* if $stockout\_sub\_withoutcall > 5$.

Controls. We now consider the control variables used in the estimation.

Worker Behaviors: We first introduce variables on worker behaviors, i.e., $OSF_{i,s}$, $CSF_{i,s}$, $worktime\_lag_{i,d}$, and $hours\_weeklag_{i,d}$. $OSF_{i,s}$ is measured as the cumulative number of orders delivered by worker *i* on the day *d* before task *t*. It captures the worker's working time targeting or inertia behavior. $CSF_{i,s}$ is measured as the cumulative earnings of worker *i* on day *d* before task *t*. It controls workers' same-day earning-targeting behavior. $worktime\_lag_{i,d}$ is measured as the working time of worker *i* on the same day (as day *d*) of the previous week. $hours\_weeklag_{i,d}$ is a proxy for workers' working habits, and represents the total working time of worker *i* on the previous week. In the robustness checks (section 5.5) we implement a 2-week lag control as well.

Worker Characteristics: We control for worker characteristics using new_worker_i, which equals one if the worker joins the platform after the beginning of our observation period.

Delivery Conditions: We control for the delivery conditions with variables $precipprob_{c,d,s}$, $humidity_{c,d,s}$, $supply_{c,d,s}$, and $demand_{c,d,s}$. In particular, $precip\_hourly_{c,d,s}$ measures the precipitation probability in the city *c*, on the day *d*, and slot *s*. We use $supply_{c,d,s}$, and $demand_{c,d,s}$ to capture the market supply and demand in slot *s* of the city *c* on date *d*.

Order Characteristics: We control for order characteristics using several variables including $experience_{i,t}$, $storefamiliarity_{i,t}$, $num\_item_{i,t}$, $call\_duration_{i,t}$, $num\_substitution_{i,t}$, $store\_size_{i,t}$, $diff\_zipcode_{i,t}$, $advancetime_{i,t}$, $pickingtime_{i,t}$, $deliverytime_{i,t}$. In particular, $experience_{i,t}$ is measured as the cumulative number of orders delivered by worker *i* before day *d*. It measures the prior experience of worker *i* before the focal day in which task *t* is performed. Next, $storefamiliarity_{i,t}$ is measured as the cumulative number of orders delivered by worker *i* from the same store as task *t*. It measures the familiarity of worker *i* to the store from which task *t* has to be picked. Among other variables, $num\_item_{i,t}$ is the number of items present in the task *t*, $num\_substitution_{i,t}$ represents the number of substitutions made in case of stockouts in task *t*, $call\_duration_{i,t}$ measures the duration of time that the worker *i* was on a call with a customer while order picking, $store\_size_{i,t}$ represents the size of the store (in 1000 square meters) from which the order was picked, $diff\_zipcode_{i,t}$ is a binary indicator of whether the store postal code and customer address postal codes were different, and is used as a proxy for distance traveled. Lastly, $advancetime_{i,t}$ is measured as the period (in minutes) between the pick start time and the delivery deadline. It represents the time a worker had in advance before the delivery deadline. The idea is that the higher the time in advance, the lower the chances of delay. We have already described $pickingtime_{i,t}$ before. Finally, $deliverytime_{i,t}$ is the period (in minutes) between the order payment time at the store and the final delivery time at the customer's location.

1.5 Econometric Model

We begin by considering an appropriate specification to examine our research questions. The objective of this work is to investigate the impact of recent (within-day) work experience on gig workers' performance. First, we consider a baseline fixed effects model specification for worker *i*'s performance on task *t*, we have:

$$Performance_{i,t} = \alpha_{0,i} + \alpha_1 OSF_{i,t} + \alpha Z_{i,t} + \tau_w + \kappa_i + \varepsilon_{i,t} - (1)$$

where $Performance_{i,t}$ is worker *i*'s performance in task *t*, $OSF_{i,t}$ is the within-day experience measure, $Z_{i,t}$ is a vector of control variables, and τ_w and κ_i are the day-of-the-week, and individual fixed effects, respectively. Note that $\varepsilon_{i,t}$ is the error term. Directly estimating model (1) via ordinary least squares regression and interpreting the estimated parameter α_1 as the effect of same-day experience on worker performance measures might lead to biased estimates due to two potential empirical challenges, namely, self-selection bias, and endogeneity (Guha, 2023), which we discuss in sections 4.1 and 4.2.

1.5.1 Correction for Self-selection Bias: Heckman Selection Model

A major empirical challenge in the gig work setting is that each worker endogenously determines whether to work, thereby leading to a self-selection bias (Guha, 2023). To address this estimation bias, we employ the Heckman selection model (Heckman, 1979), which corrects the self-selection bias through the first-stage choice equation (i.e., whether to work in a specific day and hourly slot of any given week). Furthermore, following Semykina and Wooldridge (2010), to better estimate the Heckman selection model, one needs an exogenous variable in the first stage, which affects the first stage choice equation but does not directly affect the second stage level equation (Guha, 2023). In our setting, gig workers can select the days of the week, and the hourly slots within a day in which they prefer to work. Therefore, the first stage choice equation estimates gig workers' decision on whether to work in a specific slot on a given day ($worked_{i,s,d}$), and the second stage level question estimates the worker performance on a focal task ($Performance_{i,t}$). Note that an order can start in one slot and end in the following slot. However, when it comes to determining slot selection, we consider the slot in which the picking started. We do so because our focal platform allocates tasks only in those slots in which the workers made themselves available (Guha, 2023). Following Allon et al. (2023), Xu et al. (2023), and Dai et al. (2022) we introduce a variable denoted as *avgcomp_last*, which is the average compensation per order received by the focal gig worker from the platform in her previous working day. The variable *avgcomp_last* is likely to affect workers' working decisions because such decisions (i.e., whether to work or not) mainly rely on the workers' recent work experience. This variable, however, is unlikely to directly affect the worker's performance for the focal task because the compensation amount varies across days. Since gig workers observe a different compensation amount offered by the platform once they start working on a new task on a new day, their performance is more likely to be affected by this new compensation amount (Guha, 2023). We can write the selection model as follows:

$$worked_{i,s,d} = \begin{cases} 1 \;\; if \;\; worked^{*}_{i,s,d} \; > \; 0 \\ \\ 0 \;\; otherwise \end{cases} \quad -(2)$$

$$worked^{*}{}_{i,s,d} = \beta_{0,i} + \beta_d avgcomp\_last_{i,d} + \beta X_{i,s,d} + u_{i,s,d} \quad -(3)$$

$$Performance_{i,t} = \begin{cases} Performance^{*}{}_{i,t} \;\; if \;\; worked_{i,s,d} \; = \; 1 \\ \\ unobserved \qquad otherwise \end{cases} \quad -(4)$$

$$Performance_{i,t} = \gamma_{0,i} + \gamma_1 OSF_{i,t} + \gamma Z_{i,t} + \theta \lambda_{i,s,d} + \varphi_w + \eta_i + v_{i,t} \quad -(5)$$

$$\begin{bmatrix} \sigma_u^2 \\ \sigma_v^2 \end{bmatrix} \tilde{\ } N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & \rho\sigma_v \\ \rho\sigma_v & \sigma_v^2 \end{bmatrix} \right) \quad -(6)$$

where equations (2) and (3) correspond to the first stage choice equation and equations (4) and (5) correspond to the second stage level equation.

The first stage choice equation is estimated through a Probit model where $worked^*_{i,s,d}$ is the latent variable and $u_{i,s,d}$ is the error term. We also include the exogenous variable prom $avgcomp\_last_{i,d}$ in the first stage. Other control variables contained in vector $X_{i,s,d}$ includes $CSF_{i,t}$, $precipprob_{c,d,s}$, $humidity_{c,d,s}$, $supply_{c,d,s}$, $demand_{c,d,s}$, $hours\_weeklag_{i,d}$, $worktime\_lag_{i,d}$, and new_worker_i.

The second stage equation is estimated by including the inverse Mills ratio (IMR), denoted as $\lambda_{i,s,d}$, computed through the first stage to control for selection bias. We consider three task performance measures (for $Performance_{i,t}$) - the order picking time, the delay time, and the order substitution count (when a stockout-based substitution was requested by the customer). The key variable of interest in this study is $OSF_{i,t}$, which measures the gig worker's same-day task experience computed as the number of orders delivered before the current task, *t*. Note that $OSF_{i,t}$ will be set to 0 at the beginning of each day. Only after a worker has delivered her first order, $OSF_{i,t}$ is updated to 1 (Guha, 2023). Other control variables in the level equation are included in the vector $Z_{i,t}$, that is, $experience_{i,t}$, $storefamiliarity_{i,t}$, $num\_item_{i,t}$, $num\_substitution_{i,t}$, $call\_duration_{i,t}$, $store\_size_{i,t}$, $advancetime_{i,t}$, $diff\_zipcode_{i,t}$. The variable $v_{i,t}$ is the error term in the level equation.

Following Dai et al. (2022), to alleviate the concerns of multicollinearity of the regressors in both stages and weak nonlinearity of the IMR, we use different controls in vectors $X_{i,s}$ and $Z_{i,t}$, while letting $X_{i,s}$ contain all variables in $Z_{i,t}$ that are exogenous to the sample-selection bias (Semykina and Wooldridge, 2010). In other words, $X_{i,s}$ includes all variables in $Z_{i,t}$ that are observable regardless of workers' working decisions, such as weather and past availability to work. Finally, we also introduce day-of-the-week fixed effects (φ_w), and individual fixed effects (η_i) to control for the behavioral changes across the day of the week, and unobserved individual characteristics. We present the correlation between the choice and level questions in (6), where ρ is the correlation coefficient of $u_{i,s}$ and $v_{i,t}$, and σ_v is the standard deviation of $v_{i,t}$. Lastly, due to the inclusion of IMR from the choice equation, we bootstrap the standard errors of both stages, and write a program to model the same (Semykina and Wooldridge, 2013, 2017; Guha, 2023).

1.5.2 Endogeneity Concerns and Instrumental Variables

Directly estimating gig workers' task performance by regressing the within-day hours worked on task performance ($Performance_{i,t}$) could lead to biased estimates. Our concern arises because the measure of within-day experience ($OSF_{i,t}$) is potentially endogenous as workers could be working for other organizations, either separately, or simultaneously with our focal organization. This could lead to an omitted variable bias, as learning and fatigue from same-day work in another organization, outside our focal organization, could affect the performance of the focal task (Guha, 2023). Huckman and Pisano (2006) find that task performance is not fully portable across firms, and a portion of the performance is firm-specific. As a result, we expect that within-day task experience in a different industry is not likely to affect the gig workers' focal task performance. However, same-day experience in another delivery platform could indirectly impact workers' learning and fatigue (Guha, 2023). On similar lines, we also expect that within-day compensation earned ($CSF_{i,t}$) is also subject to endogeneity as compensation earned from other platforms can affect the earnings target of the worker, thereby affecting her focal task performance. To check for possible endogeneity in our regressors, we perform the Durbin-Wu-Hausman test for endogeneity, which reveals a score of 268.31 (for picking time) and 750.48 (for delay) with a p-value of less than 0.001, rejecting the null hypothesis that our regressor is exogenous. Therefore, $OSF_{i,t}$ and $CSF_{i,t}$ in the level equation are most likely endogenous.

To account for endogeneity in the choice equation, we employ the control function probit method, owing to the non-linearity of the model (Wooldridge, 2014). In the level equations, we employ the instrumental variables (IV) approach (*xtivreg* in Stata) for the picking time and delay equations. For the substitution equation, our DV is a count variable. Therefore, we employ the control function Poisson model (Lin and Wooldridge, 2019).

A valid IV should satisfy both relevance and exclusion criteria (Wooldridge, 2010). To satisfy the relevance criterion, the IV should be correlated with the endogenous variable, and to satisfy the exclusion criterion, the IV should be independent of the error term. In other words, the IV should be correlated with the dependent variable only through the endogenous variable (Guha, 2023). We introduce the variable $OSF\_other_{i,s}$, which represents the average hours worked by other workers on the same day in the same city before the focal slot *s*. This IV satisfies both the relevance and exclusion conditions for the following reasons. For the relevance condition, $OSF\_other_{i,s}$ and $OSF_{i,t}$ are naturally correlated because of the underlying mechanism by which workers elect to work. Since the demand for, and supply of workers in a city on a given date can affect the availability of all workers, we expect that the average hours of availability of other workers before the focal slot satisfies the relevance condition. Furthermore, we run the two-stage least squares (2SLS) regressions for the OSF level equation with this IV (See Online Appendix Tables A.22, A.23, A.24 for first stage estimates of model 1). We find that the coefficient of $OSF\_other_{i,s}$ in the first-stage regression is statistically significant under a 1% significance level for all three dependent variables, which verifies the relevance condition statistically. Our instrument satisfies the exclusion restriction as well because the availability of other workers is unlikely to affect the task performance of the focal worker (Guha, 2023). Any correlation with the dependent variable(s) would only be through the endogenous variable, $OSF_{i,s}$. For the interaction equations, we need another instrument for the model to be identified. Following Wooldridge (2010), we argue that in the absence of additional instruments, we need to create additional instruments for the interaction term. We thus interact the moderators with our IV, $OSF\_other_{i,s}$ (e.g., $experience^{*}OSF\_other_{i,s}$) to identify the interaction terms. We apply a similar logic to devise an instrument for $CSF_{i,t}$ and introduce the variable $CSF\_other_{i,s}$, which represents the average compensation received thus far by other active workers in the same city-slot of the focal day.

We further compute the Kleibergen-Paap rk LM statistic (6.4*10^5) for the under-identification test with a p-value < 0.001, which rejects the null hypothesis that the equation is under-identified, that is, the IV is correlated with the endogenous variable. Next, we compute the Cragg-Donald Wald F-statistic for the weak identification test (5.2*10^5), and obtain values that are larger than the threshold value under a 10% maximal IV size (7.03) provided by Stock and Yogo (2005). These tests further support that OSF_otheri,s is not a weak instrument. Note that the difference between the two statistics mentioned above is that the Cragg-Donald Wald F-statistic assumes the error terms to be independent and identically distributed, and the Kleibergen-Paap rk Wald F-statistic omits this assumption. Specifically, the latter is relatively more robust. Finally, the Hansen J statistic for overidentification reveals that our equation(s) are exactly identified.

Lastly, in our robustness check, we consider an alternative IV for $OSF_{i,t}$. Following Allon et al. (2023); Dai et al. (2022), this alternative IV ($OSFoth\_lag_{i,t}$) is based on the notion of coworkers, where *coworkers* are defined as other active workers on the same day *d*, and slot *s* of the focal city, who made similar working decisions (i.e., whether to work) as our focal worker. More specifically, if the focal worker worked on the same day of the previous week, coworkers would be those who were available to work on that same day of the previous week. We define $OSFoth\_lag_{i,t}$ as the average number of orders delivered by other active workers working on the same day *d*, before the focal slot (in which worker *i* is working) of the previous week. This alternative IV ensures that the relevance criterion is met because the focal worker *i* and coworkers have similar (shared) working decisions in the past (Allon et al. 2023). The exclusion criterion for this alternative IV is also satisfied, as past decisions to work should not impact the performance of the focal task.

1.6 Empirical Results

1.6.1 Main Results: The Impact of Same-day Experience on Performance

We start our analyses by examining the effects of same-day picking and delivery experience (i.e., the cumulative number of orders delivered on the same day) on gig worker performance. In particular, we consider two types of performance outcomes: immediate operational outcome, i.e., picking productivity, measured using order picking time ($pickingtime_{i,t}$), and ultimate performance outcome, or service quality, measured in two different ways - First, the delay time per order ($delay_{i,t}$), and second, the number of substitutions made in an order if there is a stockout, and the customer requested a substitution ($subst\_when\_reqd_{i,t}$).

1.6.2 Effects of Same-day Experience on Picking Time

Table 1.3 presents the estimation results of our main model on the order picking time.

Column (1) presents the results with only the linear term of same-day experience (i.e., $OSF_{i,t}$), column (2) presents the results with both linear and quadratic terms (i.e., $OSF_{i,t}$ and $OSF2_{i,t}$). In addition, columns (3) to (6) consider the interaction effects of same-day experience with prior experience, order batching, task complexity, and discretion in complex tasks (i.e., $experience_{i,t}$, $num\_batched_{i,t}$, $ccomplexity_{i,t}$, $cdiscretion_{i,t}$) respectively. We find that same-day experience reduces the order picking time (i.e., improves worker's picking productivity) across all models, which is consistent with existing results on the positive same-day learning effects in the literature (Staats and Gino, 2012). We observe that the main effect is non-significant in models 1, 5, and 6, which could indicate a possible curvilinear relation. We test this by including a quadratic term (*OSF2*) in column (2), and observe that there is a non-linear (increasing with time) relationship between same-day

Table 1.3: Two-stage Heckman Selection - Picking Time

	(1)	(2)	(3)	(4)	(5)	(6)
OSF	-0.065***	-0.087***	-0.793***	-0.119***	-0.035***	-0.053***
	(0.012)	(0.019)	(0.029)	(0.012)	(0.012)	(0.012)
CSF	-0.026***	-0.027***	-0.021***	-0.026***	-0.019***	-0.026***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
experience	-0.644***	-0.644***	-0.895***	-0.595***	-0.622***	-0.651***
	(0.011)	(0.011)	(0.014)	(0.011)	(0.010)	(0.010)
storefamiliarity	-0.931***	-0.932***	-0.931***	-0.999***	-0.846***	-0.939***
	(0.008)	(0.008)	(0.008)	(0.008)	(0.008)	(0.008)
IMR	0.644***	0.617***	0.948***	0.666***	0.489**	0.679***
	(0.236)	(0.237)	(0.236)	(0.236)	(0.234)	(0.233)
OSF2		0.003				
		(0.002)				
OSF_exp			0.115***			
			(0.004)			
OSF_cbatch				0.095***		
				(0.005)		
num_batched				0.569***		
				(0.016)		
OSF_ccomplex					-0.059***	
					(0.002)	
ccomplexity					1.010***	
					(0.006)	
OSF_cdiscretion						-0.017***
						(0.003)
cdiscretion						-1.007***
						(0.008)
Observations	1,661,655	1,661,655	1,661,655	1,661,655	1,661,655	1,661,655
Picker fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Day fixed effect	Yes	Yes	Yes	Yes	Yes	Yes

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

orders delivered and picking time. We further observe the main effect of *OSF* on picking time is now negative and significant. We also test the same for models 5 & 6 in the Online Appendix, Table A.20. Next, in column (3), we find that the interaction term of same-day experience and prior experience is positive and significant, although the main effect of *OSF* and *experience* on picking time is negative and significant. This could indicate a complementary relation between the two variables. In column (4), we find that batch picking increases picking time. Results from the interaction demonstrate that as within-day experience increases, the higher the number of batched orders, the higher the picking time per order. We perform a robustness check (Online Appendix, Table A.4, Col. 4) using a binary (1/0) measure and find similar results. In column (5), we find that higher complexity increases picking time. We further find the interaction term of same-day experience and complexity to be negative and significant, which implies that productivity increases (picking time reduces) when gig workers with higher within-day experience handle complex tasks. Furthermore, since our task allocation algorithm relies on the simplicity of the allocations, we classify complexity into 3 levels (high, medium, low) based on the number of stockout items for which substitutions were requested, and conduct a robustness check to validate our results and find a similar conclusion. Finally, in column (6), we find that discretion while performing complex tasks reduces picking time, thereby increasing productivity. We further find that workers with higher same-day experience are more productive while handling tasks requiring discretion. Interaction plots are displayed in Figure A.3

1.6.3 Effects of Same-day Experience on Delay

Next, we discuss the effects of same-day delivery experience on delay time (See Table 1.4). Surprisingly, we find that same-day experience increases order delay across all models. It could be possible that within-day experience could be detrimental to service quality, when within-day experience is low (Dai et al., 2022). To examine this effect, we test the presence of a non-linear relation in model 2 using the quadratic term $OSF2_{i,t}$. We also include the quadratic term in the remaining models 3 & 4 in the Appendix, Table A.21. We observe the presence of a non-linear relationship in column 2, which indicates that higher within-day experience can reduce order delays.

Table 1.4: Two-stage Heckman Selection - Delay

	(1)	(2)	(3)	(4)
OSF	0.184***	0.825***	1.516***	0.202***
	(0.016)	(0.027)	(0.040)	(0.017)
CSF	0.044***	0.048***	0.033***	0.044***
	(0.002)	(0.002)	(0.002)	(0.002)
experience	-0.044***	-0.031***	0.414***	-0.055***
	(0.011)	(0.011)	(0.017)	(0.011)
IMR	-1.659***	-0.875***	-2.209***	-1.628***
	(0.326)	(0.327)	(0.326)	(0.326)
OSF2		-0.098***		
		(0.003)		
OSF_exp			-0.209***	
			(0.006)	
OSF_cbatch				-0.052***
				(0.007)
num_batched				1.037***
				(0.022)
Observations	1,661,655	1,661,655	1,661,655	1,661,655
Picker fixed effects	Yes	Yes	Yes	Yes
Day fixed effect	Yes	Yes	Yes	Yes

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

In column (4), contrary to our expectations, we find that batch picking increases delays as same-day experience increases. This finding, however, is not surprising. Since in our setting, the picking and delivery is performed by the same worker, unlike (Dai et al., 2022), where workers only perform deliveries. In such cases, batching can add to the complexity of the task, especially if the number of batched orders is high. Furthermore, search time increases while picking batched orders due to different items, and different order arrival times, thereby increasing order picking time, which can further contribute to order delays. We perform a robustness check using a binary measure of order batching in Table A.5 of

Appendix and find similar results. Interaction plots displayed in Fig. A.4

1.6.4 Effect of Same-day Experience on Items Substituted when Requested

Table 1.5 shows the effect of the same-day experience on stockout-based substitutions.

Table 1.5: Two-stage Heckman Selection - Substituted when requested

	(1)	(2)	(3)	(4)	(5)	(6)
OSF	0.030***	0.029***	0.023***	0.030***	-0.004***	0.024***
	(0.001)	(0.001)	(0.002)	(0.001)	(0.001)	(0.001)
CSF	-0.001***	-0.001***	-0.001***	-0.001***	-0.000	-0.001***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
experience	0.104***	0.104***	0.102***	0.103***	0.041***	0.077***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
storefamiliarity	-0.154***	-0.154***	-0.154***	-0.152***	-0.023***	-0.109***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
IMR	-0.232***	-0.222***	-0.229***	-0.233***	-0.218***	-0.206***
	(0.019)	(0.019)	(0.019)	(0.019)	(0.019)	(0.019)
IMR2	-1.018***	-1.022***	-1.018***	-1.015***	-0.594***	-0.951***
	(0.010)	(0.010)	(0.010)	(0.010)	(0.010)	(0.010)
OSF*OSF		-0.000				
		(0.000)				
OSF*experience			0.001***			
			(0.000)			
num_batched				-0.022***		
				(0.001)		
OSF*num_batched				0.002***		
				(0.000)		
ccomplexity					0.185***	
					(0.000)	
OSF*ccomplexity					-0.001***	
					(0.000)	
cdiscretion						0.143***
						(0.000)
OSF*cdiscretion						-0.001***
						(0.000)
Observations	1,275,549	1,275,549	1,275,549	1,275,549	1,275,549	1,275,549
Picker fixed effects	yes	yes	yes	yes	yes	yes
Day fixed effects	yes	yes	yes	yes	yes	yes

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

As expected, a higher same-say experience increases item substitutions. In column

(2) we test for the presence of a non-linear relationship between same-day hours worked and count of substitutions. The effect remains positive and significant. In column (4), we find the interaction term for order batching and same-day experience is negative and non-significant. We conduct a robustness check with a binary measure of batching in Online Appendix, Table A.6, to find a negative and significant interaction, which indicates that substitutions decreases as workers' within-day experience increases, indicating a possible lack of attention to detail. In column (5), we find that the interaction of within-day experience and complexity is negative and significant, which indicates that as within-day experience increases, there are fewer substitutions in more complex tasks. Finally, in column (6), we find that the interaction term of within-day experience and complexity is negative and significant, which indicates that as within-day experience increases, there are fewer substitutions in tasks where workers use their discretion. Robustness checks using a categorical measure (low, medium, high) of complexity and discretion in Online Appendix, Table A.6 indicate that within-day experience increases substitutions at moderate levels of complexity and discretion, while it decreases substitutions when complexity or discretion is high. Interaction plots are displayed in Figure A.5. Additionally, plots testing for the presence of a non-linear relationship are displayed in Figure A.6

Table 1.6 presents a summary of the hypothesized results.

Table 1.6: Support for Hypotheses tested

	(1) pickingtime	(2) delay	(3) substituted_when_reqd
H1 (OSF)	✓	✓/X	✓
H2 (OSF*batching)	✓/X	X	✓
H3 (OSF*complexity)	✓	-	✓
H4 (OSF*discretion)	✓	-	✓

✓ *fullysupported*, X *notsupported*, ✓/X *partiallysupported*

1.6.5 Robustness Checks

We perform robustness checks to test the sensitivity of our results, and to rule out plausible alternative explanations. Table 1.7 provides a summary of the tests performed. In particular, we consider alternative measures, alternative samples, and alternative specifications.

Table 1.7: List of Robustness Checks

Test Type	Description	Table
Alternative independent variable	Same-day hours worked so far ($HSF_{i,t}$)	Tables A.7,A.8,A.9
Alternative moderating variables	a)Presence of batch picking (0/1)	Tables A.4,A.5,A.6
	b)Complexity levels (Low, Medium, High)	Tables A.4,A.5,A.6
	c)Discretion levels (Low, Medium, High)	Tables A.4,A.5,A.6
Alternative instrument (IV)	Orders delivered by coworkers on the same day of last week ($OSF\_lagother_{i,t}$)	Tables A.10,A.11,A.12
Alternative samples	a)Subset of workers with sample average delivery experience above mean	Tables A.13,A.14
	b)Only workers with average working hours in top 25% of the platform	Table A.15
	c)Workers with minimum delivery experience of 4 orders or more (bottom 10% removed)	Table A.15
Alternative Specifications		
a) Alternative measure of behavior	Including a 2-week lag of $hours\_weeklag$	Table A.18
b) Alternative measure of weather	Excluding $humidity_t$ and $precip\_prob_t$ and including $precip\_hourly_t$	Table A.18
c) Additional control for peak hours	Additional variable to indicate peak hours	Table A.19

Alternative Measures: In this section, we discuss the robustness checks with alternative measures. First, we use an alternative independent variable. In our main analyses, we use the cumulative number of orders delivered within the same day ($OSF_{i,t}$) as our key measure for the same-day learning experience. In the robustness check, following (Xu et al., 2023; Chan et al., 2014; Dai et al., 2022), we consider an alternative measure computed as the cumulative hourly slots worked before the focal slot on the day *d* ($HSF_{i,t}$). Moreover, we also replace the corresponding instrument to $HSF\_other_{i,t}$, which is the average hourly slots worked by other active workers on day *d*, in the same city, before the focal slot. The estimation results are shown in Tables A.7, A.8, and A.9 of the Online Appendix, and are consistent with our main results.

We also use alternative measures for the moderators. Instead of continuous measures for batching, complexity, and discretion, we use binary or categorical measures of the variables. Refer to the variable descriptions in section 3.2. The results, presented in the Online Appendix Tables A.4, A.5, and A.6, are in-line with our main estimations apart from the coefficient of OSF in models 5 & 6 of Tables 4 (Picking time) and 5 (Delay). We further note that the coefficients of *med_complexity*, *high_complexity*, *med_discretion*, *high_discretion* allow us to obtain more nuanced understanding of when the effect of *OSF* on *pickingtime*, *delay*, and *substituted_when_reqd* is most pronounced.

Next, we use an alternative IV, $OSF\_lagother_{i,t}$ to perform the estimations. Refer to section 4.2 for the variable description. The results, presented in the Online Appendix Tables A.10, A.11, and A.12, are mostly in-line (both in direction and significance) with our main estimations. Although the direction (and the magnitude) of the $OSF_{i,t}$ coefficient(s) is as expected, they are not significant. Therefore, to further validate the effect of $OSF_{i,t}$ on gig workers' performance, we conduct additional tests with alternative samples to rule out alternate explanations.

Alternative Samples: First, to alleviate the potential endogeneity concern that workers with good performance outcomes may also be those who ultimately obtain higher experience, we conduct robustness checks following (Xu et al., 2023) and (Guha, 2023), with a subset of workers who achieve the sample average delivery experience of at least 266 orders. We present the results in the Online Appendix Table A.13. The dependent variable (DV) of model 1 is picking time, that of model 2 is delay, and model 3 is stockout substitutions when requested by the customer. Models 1, 2, and 3 of the table represents m1 of our main regressions presented in Table 1.3, 1.4, and 1.5 respectively. We observe that the coefficient of $OSF_{i,t}$ is consistent with our main models. Although, we do observe that the coefficient is non-significant in model 1. One explanation could be that since we are selecting workers with higher experience the effect of *OSF2* will be more prominent.

To further examine this effect, we take a sub-sample of workers who achieve the sample average delivery experience of 1 standard deviation above the mean, i.e., at least 702 orders. We present the results in Table A.14. The coefficient of $OSF2_{i,t}$ is in-line with that of our analyses.

Next, workers in our estimation sample might take orders from other platforms, which we cannot observe. Although we use two different instruments to address this issue, we conduct some additional analysis to further reduce this concern. First, we run the regression with a sub-sample of workers actively involved in our focal platform. In particular, we consider workers with average working hours of the top 25% in our focal platform. We present the new results in the Online Appendix Table A.15, and find that the coefficients (of $OSF2_{i,t}$ and other variables) are consistent with our main analyses. Next, we conduct another analysis by taking a sub-sample of workers after removing bottom 10% inactive workers, i.e., those workers who have delivered less than 4 orders during their tenure. Again, we find that the coefficients are consistent with our main analyses.

Alternative Specifications: Next, we discuss checks with alternative specifications. First, in our main estimation, we proxy for workers' working habits with the total working time in the previous week ($hours\_weeklag_{i,d}$). However, workers' working habits may last more than just one week (Xu et al., 2023). Therefore, we add controls for an extended lag period of two weeks ($hours\_2weeklag_{i,d}$) to better capture working habits. We present the estimations in Table A.17 of the Online Appendix and find that the results are consistent with our main analyses.

Next, we include an alternative measure for weather. Instead of the humidity and precipitation probability in an hourly slot, we include ($precip\_hourly_{i,s}$), which is the actual hourly precipitation in each slot. The estimation results, presented in Table A.18 of the Online Appendix, are in-line with our main analyses. Although, the coefficient of *OSF* for *pickingtime* is not significant.

Finally, in our main analyses, we do not include peak hour dummies as the hourly slot selection already factors in the demand and supply of each slot, and hence the effect of *peak hours*. However, following Dai et al. (2022), we include a control for peak hours, which in our dataset correspond to 8 a.m. to 7 p.m. The hours before and after this period are non-peak hours. We present the estimation results in Table A.19 of the Online Appendix and find that the results are consistent with our main analyses.

1.7 Task Allocation Algorithm

In this section, we seek to understand the performance implications of study. Specifically, the effect of within-day experience (*OSF*) and prior task experience (*experience*). The existing policy specifies that all available workers in a city, on a given date, and time slot should be allocated work, depending on the number of orders in that *city-date-slot*. If the number of gig workers is greater than the number of orders, then some gig workers are not allocated tasks. On the other hand, if the number of orders is greater, some gig workers might need to deliver more than one order in that slot. However, there is no defined way of allocation, which means an available worker could be randomly allocated in any order. If the order is not accepted by that worker, then it would be allocated to a new worker, and so on (Guha, 2023).

Our objective is to devise a simple assignment problem where there are a set of tasks (T), a set of workers (W), and a set of constraints (C). The objective of our algorithm is to assign tasks based on one simple criterion - Allocate more complex tasks to higher *ranked* workers. However, the *greedy* policy of simply prioritizing higher ranked workers to more challenging tasks is often inefficient, especially when it is not bounded by constraints. We therefore allocate workers to tasks based on three broad levels of task complexity, while restricting the maximum number of orders a worker can deliver (batch) in an hour, so that the aggregate performance across all workers improves.

The level of complexity (high, moderate, or low) is determined using the variable, *complexity*). To calculate rank, we utilize the variable orders delivered so far ($OSF_{i,t}$) along with the variable $experience_{i,t}$ (prior experience). We first standardize all variables, excluding the binary and categorical variables, and run estimations on the three outcome variables (*pickingtime, delay, substituted_when_reqd*). We run the estimations on the main model(s), excluding squared and interaction terms. The absolute value of the standardized coefficients of *OSF*, and *experience,* are used to rank the predictors. The variable that has a higher importance has a higher absolute value. We determine the ratio of the coefficients in each of the three estimations (Guha, 2023). In each estimation, the coefficient of *experience* is twice (or more) than that of *OSF*. Therefore, we develop the rank of the worker based on the following equation:

$$rank_{i,t} = OSF_{i,t} + 2 * experience_{i,t} \text{ - (7)}$$

We run our algorithm using the entire dataset. Same-day experience and prior experience are calculated dynamically as new allocations are made each day. While making the allocations, we impose multiple constraints to reflect the realities of our setting closely. These include taking into account the hourly availability of the shoppers within a day. We further ensure that no individual worker conducts more than two tasks in a given hour, because orders are more likely to be delayed if the same worker picks more than two orders in an hourly slot, especially if the order deadline is close (Guha, 2023). This constraint further ensures that workers are not fatigued by the frequency of orders, and helps maintain better productivity and service quality. Other constraints include allocating workers within the same city in which they operate. Overall, the objective of the algorithm is to sort workers based on their rank (from highest to lowest), and then allocate them to tasks based on task *complexity*. Once the new allocations are made by the algorithm, we calculate the predicted *pickingtime*, *delay*, and *substituted_when_reqd* based on the newly allocated workers' *experience* and *OSF* (Guha, 2023). To make the predictions, we run our main analyses, including the quadratic term of OSF ($OSF2_{i,t}$). Furthermore, following Avgerinos and Gokpinar (2017), we hold all task-specific variables (for e.g., number of substitutions, call duration, picking time, etc.) at their mean. Workers' task-specific variables (such as, experience, store familiarity, etc.) are calculated dynamically with each new allocation, and included as controls in the choice and level equations. We further recreate all variables that denote worker behaviors (for e.g., working time in past week, compensation so far, etc.) and include them in both stages of the new selection equation. The level equation is denoted as: (Refer to section 4.1 for the notations)

$$Performance_{i,t} =$$
$$\gamma_{0,i} + \gamma_1 OSF_{i,t} + \gamma_2 OSF2_{i,t} + \gamma_3 experience_{i,t} + \gamma Z_{i,t} + \theta \lambda_{i,s,d} + \varphi_w + \eta_i + v_{i,t} \quad -(8)$$

We then compare the improvements in productivity and service quality from the proposed task allocation rules based on our econometric model. Table 1.8 below presents the results of the percentage change in picking time, delay, and stockout-based substitutions, upon comparison with the baseline *random allocation model*, which is currently being followed by the platform.

Table 1.8: Task Allocation Algorithm - Baseline Comparison

	(1) Mean	(2) 75th percentile
Percentage of change in *picking time*	-6.90%	-7.58%
Percentage of change in *delay*	48.40%	-6.26%
Percentage of change in *cv_delay*	-75.76%	-
Percentage of change in *substituted_when_reqd*	28.85%	22.42%

Comparison of the algorithms. As expected, our algorithm (based on complex task assignment to higher ranked workers) results in better productivity and service quality compared to the current allocation method. In column 1 we present the comparison of the average (mean) values and in column 2, we present the comparison of the 75^{th} percentile values. The 75^{th} is selected because in the top quartile, the algorithmic (out-of-sample) predictions are expected to perform better than the in-sample predictions.

Surprisingly, we note that mean *delay* increases by 48% points when we use our algorithm. A reason for this could be that important controls such as picking time, delivery time etc. are unobserved for the algorithmic prediction. However, we observe that at the 75^{th} percentile, delays reduce by 6% points by using the algorithm for allocations. Following

Dai et al. (2022), we also calculate the percentage change in *cv_delay*, i.e., the coefficient of variance (CV) in delay. *cv_delay* is a measure of volatility (or consistency) in delay time. In a recent study, Bavafa and Jónasson (2021) demonstrate that experience affects the variability in outcomes, which is an important measure of quality. They further assert that ignoring the consistency measure might underestimate the benefits of learning. We observe a 76% reduction in delay variability when workers are systematically allocated to tasks based on a set of rules.

To demonstrate the practical application of our work, we devised a method where the algorithm will self-refresh in a certain duration of time and then assign the remaining new tasks to the available workers. This delay ensures that (highly-ranked) workers do not get allocated to any available task (of low complexity). By slightly prolonging this delay, we ensure that there are sufficient new tasks that would make the allocation process better. A recent work by Xin and Xie (2023) examines how slightly delaying real-time decisions can help the decision-maker gather more information to make better decisions. That said, it would make sense to learn more about how the success of the algorithm depends on different parameters implemented by similar on-demand platforms in other countries. Managerial insights regarding where our policy might be most promising could be an interesting area of future research.

1.8 Extension: A Minimalist Model for Batched Orders

In this section, we describe our ongoing work, which we will later incorporate as part of the theory and model building sections.

1.8.1 Single Store, Single Worker Model for Batched Orders

Let us consider a setting with a single grocery store, and a single gig worker, initially. Let each order (o_i) of size n_i items have a processing time of $p(n_i)$. The first order is received

at time t_i and has a promised a delivery time of T_i.
This implies, the promised delivery deadline, $D_i = t_i + T_i$; and actual delivery time is A_i

Since, our setting uses a dispatch mechanism, jobs (j) are released to the workers, where each job j, is of size b. If the orders are not batched, batch size, b = 1 $\rightarrow$ i and j are interchangeable.

If b $>$ 1, orders have been consolidated into batches with index j for batch. Let r_j be the release time of job j and I_j be the dispatch delay before the job is accepted by a worker w (for b=1, we assume $r_j = t_i$).

Let N_j be the inbound travel time to the store (for simplicity, we assume = 0), and let $P_j(b)$ be the consolidated picking time for the b batched jobs.

The worker will reach the store at: $t_i + I_j + N_j$

The worker will complete picking at the store at $t_i + I_j + N_j + P_j(b)$

The worker will complete job, j by arriving at customer i at time $A_i = t_i + I_j + N_j + P_j(b) + O_j$ where, O_j is the outbound travel time to the first customer of job j

$\triangle_{k,k+1}$ is the incremental travel time from customer k to customer k+1 of job j

If A_i (actual) $\leq D_i$ (deadline), i.e., $I_j + N_j + P_j(b) + O_j \leq Ti \rightarrow$ no delay

The flowtime (F_j) of order i, $F_i = I_j + N_j + P_j(b) + O_j$;

Labor time of job j, $L_j = N_j + P_j(b) + O_j$ at unit cost c_w; and

Travel time for job j, $T_j = N_j + O_j$ at unit cost c_t

1.8.2 Effect on Delays in Batched Orders

Assuming a homogeneous and deterministic processing time P(1) for each order, we have, for jobs of size b = 1:

$F_1(1) = I + N + P(1) + O$; Labor per order = $N + P(b) + O$; Travel per order = $N + O$

For batched orders (say, b = 2), let the inter-arrival time (of the orders) be τ.

If the job is released at time I, the flowtime for order 1 of 2:

$F_1(2) = \tau + l + N + P(2) + O$ vs $F_1(1) = l + N + P(1) + O$

Therefore, $F1(2) - F1(1)$ = $\tau + [P(2) - P(1)] > 0$ since $P(2) - P(1) \geq 0$, since picking of two orders cannot take less time than one order.

The flowtime for order 2 of 2, $F_2(2) = \tau + I + N + P(2) + O + \triangle_{1,2}$, and by parallelism to the single order per job case, $F_2(2) - F_1(1) = [P(2) - P(1)] + \triangle_{1,2} > 0$

So, each order is completed later that when the batch size is 1, so likelihood of delay can only go up unless promised lead time is made longer

1.8.3 Improvements in Operating Costs for Batched Orders

Travel time for job of size 2 = $N + O + \triangle_{1,2}$ or, $[N + O + \triangle_{1,2}]/2$ per order

Savings in travel time/order = $[N + O] - [N + O + \triangle_{1,2}]/2$ = $N + [O - \triangle_{1,2}]/2$; which is $>$ 0 since the term in [] is non-negative when travel costs go down

Labor time for job of size 2 = $N + P(2) + O + \triangle_{1,2}$ or $[N + P(2) + O + \triangle_{1,2}]/2$ per order

Labor Savings per order:

$[N + P(1) + O] - [N + P(2) + O + \triangle_{1,2}]/2$ = $N + [O - \triangle_{1,2}] + [2P(1) - P(2)]/2 > 0$ since each term is [] is non-negative because of economies of scale.

1.8.4 Learning and Forgetting Effect

For each worker k, the unit marginal cost is $C_k(Q_{j,k})$ = α(k, missed days)$\beta_k e^{-\gamma(j,k)}$, where $Q_{j,k}$ denotes the job number for worker k; and β represents the base pick rate (adjusted for job complexity and batching), and γ is the basic learning rate.

If there was only one gig worker, the earliest we can finish a task is = $1 - e^{-(\alpha_k)(\beta)Q_{t-1}}$

Extension (Ongoing work): The effect of learning and forgetting on batch picking

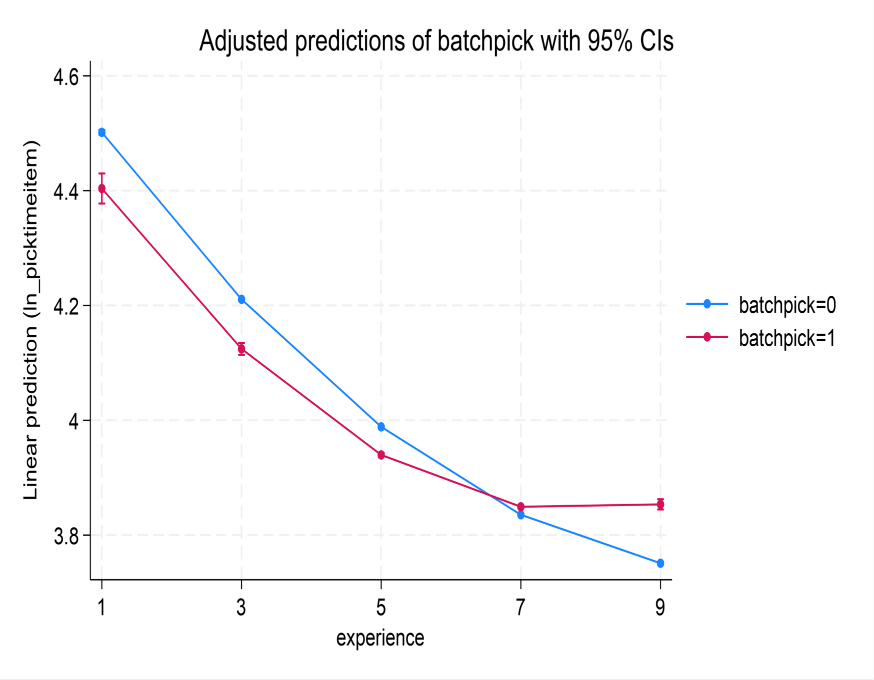
In this section, we describe our ongoing work, which we will later incorporate as part of the theory and literature sections. The idea is to examine how gig worker's learning and forgetting behaviors can affect task productivity. We further examine how task batching, task complexity, and discretion in complex tasks interact with learning and forgetting behaviors of gig workers, and in turn affects worker productivity. Figure 1.1 displays plots from a preliminary analysis, where we examine how job-based learning, and service gaps (forgetting) affect productivity (in terms of picking time per item) when the tasks involve order batching vs when it does not.

A quick glance at the plots reveal that task batching can help reduce picking time per item (i.e., improve productivity) when gig workers' experience (task-based learning) improves. On the other hand, discontinuity in service could lead to forgetting, which in turn increases the picking time per item. While batch-picked order still take less time to complete, the picking time per item increases for both batched and non-batched orders.

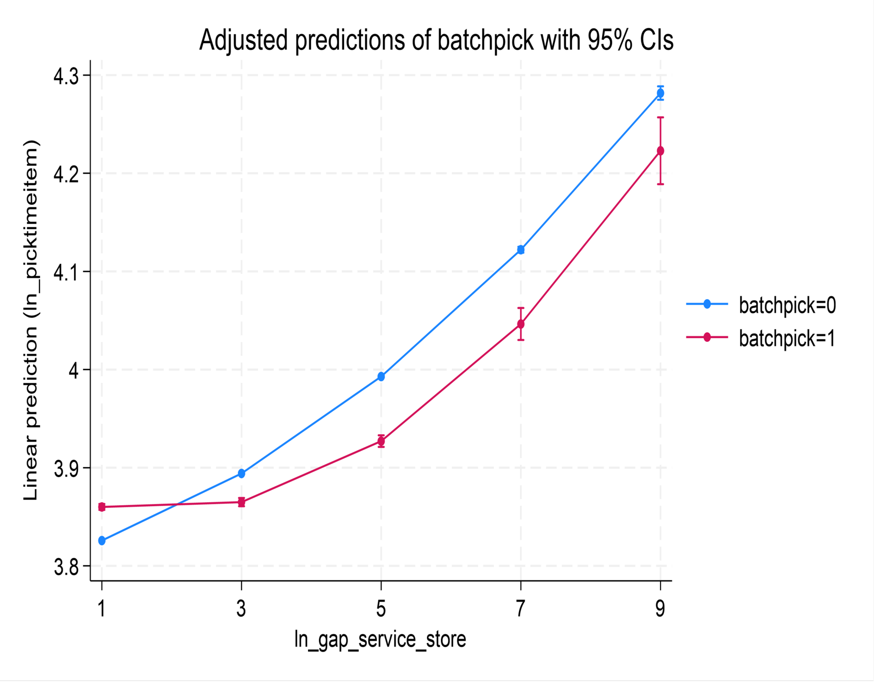
We plan to conduct this analysis for all performance variables (picking time, delays, and substitutions when requested) along with interaction effects (with order batching, task complexity, and worker discretion in complex tasks) and reframe our paper (including the theory and hypotheses) as a job (and task-specific) learning and forgetting paper. Specifically, our analysis will demonstrate that individual 'forgetting' behavior exists in gig work, which is why recent (or, intra-day) experience can be beneficial.

1.9 Discussion and Conclusion

This paper investigates how within-day and prior experience affect gig workers' performance in on-demand platforms. We further examine the moderating effects of task characteristics, such as, task batching, complexity, and worker discretion in complex tasks. Using data from a leading online grocery platform in Europe, we estimate the effects using a two-stage Heckman model with instrumental variables (IVs). We find that within-day work experience improves gig workers' productivity, and service quality. Specifically, the order picking time reduces, and the stockout item substitutions when requested by a customer, increases. Surprisingly, we find that order delays increase as same-day experience



(a) Learning*Batch Pick vs Pick time per item



(b) Forgetting*Batch Pick vs Pick time per item

Figure 1.1: Interaction effects of Learning & Forgetting on Picking time per item

increases, when same-day experience is low. However, as workers gain more delivery experience within the day, order delays reduce. We also investigate the moderating effects of order batching, task complexity, and worker discretion in complex tasks, and provide insights regarding when the benefits of longer same-day work on task performance are most pronounced. Finally, we develop a task allocation algorithm, where higher-ranked workers are assigned to more complex tasks. We find that allocating workers to tasks based on task complexity, while limiting the number of orders that can be batched, leads to improvements in gig workers' productivity and service quality.

1.9.1 Managerial Implications

Although gig-work platforms have gained massive traction in food and grocery retail, enough attention has not been paid towards how to best manage gig workers in a way such that the task performance is maximized, but not at the expense of gig workers' well-being. We first demonstrate that providing gig workers the flexibility to select hourly working slots could increase workers' task performance as workers gain from within-day experience, while minimizing the detrimental effects of fatigue and demotivation, resulting from the inability to select their work hours (Guha, 2023). In addition to improved substitutions, a better task allocation mechanism can reduce picking time per order, and order delays. In the quick-commerce business, *"time is money"*. Assuming an average picking time of 20 minutes, by simply reducing the picking time and delay by 2.5 minutes per order, a worker can save enough time to deliver an additional order after every 8 orders. This not only improves the earnings rate of the worker, but also increases overall revenue of the on-demand platform.

Our algorithm demonstrates how properly allocating gig workers to specific tasks could increase both productivity and service quality. Specifically, we demonstrate that allocating workers with a mix of higher prior experience, and within-day experience to complex tasks can have significant performance benefits, both in terms of productivity and service quality.

1.9.2 Limitations and Scope for Future Research

As with all studies with endogenous covariates, our variables are also affected by endogeneity concerns. While we have tried to address it to the best of our abilities, there could be questions regarding how specific levels of experience might interact with an individual worker's same-day experience, familiarity, or fatigue, and how that might affect productivity, and service quality. Also, since the shoppers could be working for other organizations simultaneously, addressing concerns about how experience, familiarity, and fatigue resulting from unobserved work could affect performance outcomes requires additional thought. Although our robustness checks alleviate these concerns to a certain extent, future studies could rely on more robust techniques such as field experiments to address these concerns. Additionally, using other measures of service quality, such as customer satisfaction (NPS) scores, could provide future research directions.

Lastly, we mention the generalizability of our findings. and scope for future work. Although our findings are from a single platform, we expect the implications of our study to be generalizable to other gig platforms where workers have the autonomy to select their work hours and tasks. Moving forward, research on how other elements of the gig economy, such as the store and the customer, can impact the performance and survival of these platforms would deserve attention from scholars. Another promising research direction could be examining the *substitution effect*, or a cost-benefit analysis of incorporating substitutions in on-demand grocery. We hope our findings and suggestions help provide future research directions in this area.

Chapter 2

Improving Picking Efficiency of Online Grocery Retailers When Order Pickers and Customers Co-exist

2.1 Abstract

Gig workers of on-demand grocery platforms frequently encounter the presence of customers and other in-store employees while picking orders from ‘busy’ retail stores. This affects their picking efficiency due to customer interference, and longer checkout queues; and the quality of picking due to out-of-stocks and picking errors. Although minimizing customer encounters and improving picking time could be possible through better route optimization and AI-enabled product navigation, it is often infeasible to collect store layout information from retail stores not owned by the platforms. In this paper, we investigate different techniques through which personal shoppers of online grocery platforms can improve their picking productivity and service quality, by avoiding customer encounters within retail stores. For our first approach, we schedule the picking of non-urgent orders to less-busy periods. We match similar order baskets picked during busy vs less-busy

periods to establish the value of picking during less-busy periods. Next, we develop a scheduling algorithm to schedule order picking based on urgency and order-specific characteristics. We then compare the predicted improvements in terms of reduction in order picking time and out-of-stocks. Next, we examine the benefits of using self-checkout counters and *scan-on-the-go* apps to reduce queuing time, thereby reducing the shopping time. Lastly, we utilize the results from an A/B test to determine whether adopting a separate picker and deliverer approach can help reduce order delivery time and delays. Our findings have significant implications for managers of ultrafast delivery platforms. Quick-commerce grocery deliveries are struggling to become economically sustainable, and one of the reasons is the high expense required to manage and maintain company-owned *dark stores*. Our suggestions can help platforms achieve quicker turnaround times, while maintaining the desired levels of service quality, by adopting in-store picking without investing in dark stores.

2.2 Introduction

The year 2015 saw a change in the way online grocery delivery was perceived by customers. Getir, a Turkish online grocer, promised deliveries under 10 minutes. The underlying key to this promise was the presence of a set of micro fulfillment centers (MFCs), commonly known as '*dark stores*', owned by Getir. The idea of the dark store was to offer a limited range of SKUs, that would be easier to manage, and more efficient for order pickers to pick from, without the presence of customers, thereby ensuring quicker turnarounds and better on-shelf-availability (Cheng, 2021b). This trend gained traction by 2018, when online grocers such as Migros, Glovo, JOKR, and Gorillas launched their '*quick commerce*'-based business models and promised ultra-fast deliveries from company-owned dark stores (Moreno and Gamze, 2020). Owning a set of MFCs involves additional costs in the form of rent payments, inventory management, and salary owed to part-time or full-time in-store employees. Although this additional cost burden can be transferred partly to customers by charging a ‘premium’ for quick deliveries, increased competition has forced platforms to keep their product and delivery prices low, and comparable to market prices. As a result, a major part of these additional costs are borne by the platforms. This, in turn, results in reduced profitability, and a non-sustainable business strategy for ultra-fast grocery delivery platforms (Mohamadi et al., 2023). Therefore, a question of relevance for online grocery platforms is how to sustain growth and profitability while maintaining quicker turnaround times and better service quality. Specifically, can online grocers adopt certain strategies that allow them to improve productivity and quality of service without incurring additional costs?

During the last year, owing to low operating margins, and unsustainable delivery time promises, several on-demand, ultra-fast grocery platforms either shut their operations down (Krantz, 2022; Lomas, 2022; O’Brien, 2022) or switched their operating model to include in-store pick-up options (Fickenscher and Wayt, 2022). Platforms such as Glovo started operating in a hybrid model, which included both retail store picking and ultra-fast deliveries. In the near future, as more online grocery platforms start realizing that order picking from retail stores not only allows them to offer a wider range of products, but also reduces their fixed and overhead costs, there could be a gradual shift towards retail store pickups. In addition to the online platforms, retail stores have also introduced the *buy online pickup in store* (BOPS) feature, which allows customers to place orders online, which are picked by in-store employees and later collected by the customers. With the growth in online grocery purchases, picking all items ordered online, from the backroom warehouse of stores is infeasible. Furthermore, this could mean that the number of in-store and third-party (platform) pickers inside retail stores would increase significantly in the near future (Neves-Moreira and Amorim, 2022), which in turn could reduce picking productivity. Following this, a recent branch of literature has attempted to develop better in-store picker routing algorithms to reduce customer encounters (Neves-Moreira and

Amorim, 2022) and to test the effectiveness of such AI-based algorithms on picker productivity (Knight et al., 2023). To the best of our knowledge, Knight et al. (2023) is the only study where the authors partnered with an on-demand grocery platform (Instacart) and utilized algorithms to better route pickers and locate products on retail store shelves. Using a field experiment, the authors demonstrate that there are limits to the productivity benefits of AI-enabled guidance systems, and that the benefits are most prominent when *less experienced* workers use it. Although beneficial, implementing an AI-enabled information system in an on-demand platform app where picking is performed across multiple stores and store brands is a tedious task. Therefore, in this study, we investigate alternative ways to deal with the challenges of in-store picking by examining temporal allocation of orders, and other operational strategies.

Micro fulfillment centers (dark stores) allow in-store pickers to perform order picking without customer interference (Cheng, 2021b), from a limited range of SKUs, that minimizes search and travel costs, thereby reducing picking time (Batt and Gallino, 2019). On the other hand, order picking from retail stores requires pickers to travel longer distances and encounter customer interference while picking and check-out. One way to avoid frequent customer encounters, and reduce overall shopping time while picking orders from retail stores, could be scheduling of orders to *less-busy periods*, characterized by lower store traffic. Most on-demand grocery platforms focus on improving productivity through quicker deliveries of urgent orders. However, an often overlooked aspect is the delivery of '*scheduled orders*'. Data from a European online grocery platform delivering both urgent and scheduled orders, revealed that roughly 25% of all deliveries were '*urgent*' (to be delivered immediately), while the remaining 75% were '*scheduled*' (delivered at a pre-specified time). The data further revealed that although personal shoppers had the choice to shop for pre-scheduled orders at their convenience, in reality, they pick the orders close to the delivery deadline. If the *scheduled* delivery deadlines coincide with typical '*busy hours*' in stores, personal shoppers have to pick orders alongside several other customers and in-store pickers, which not only increases the search time for items, but also the queuing time. Our idea is, within the same day, there are periods when a *store* resembles a *warehouse*, that is, there is less store traffic, which allows personal shoppers to avoid customer interference and longer queues. This could reduce the time required to shop and translate to additional orders, thereby improving productivity.

However, an alternative explanation could be that *less busy* store hours typically coincide with late evening or mid-afternoon hours, during which several store shelves (especially fresh produce) are empty. Stores often perform replenishment during these hours, owing to lesser store traffic. This, in turn, could translate to higher stockouts per order, thereby affecting service quality. In terms of productivity, although the order picking time might be shorter due to fewer customers, the overall shopping time might be similar because the queuing time in non-busy stores is likely to be longer due to fewer number of open checkout counters. The combined effect could translate to minimal or no improvement in shopping time, and higher stockouts. Furthermore, whether it is realistic to recommend picking of scheduled orders ahead of time, during less-busy hours, deserves attention. Grocery orders have ambient, chilled, and frozen items. If such items are picked too early, there could be quality issues around maintaining the cold chain and keeping the products at the desired temperatures.

Our first objective, therefore, is to determine the benefits (if any) of order picking during *less-busy* store hours. To answer this question, we first formulate an appropriate measure of *less-busy* store hours. Next, following Moreno and Terwiesch (2015) and Bell et al. (2020), we perform a matching estimation to match similar orders picked during busy (non-treated) and less-busy (treated) hours, and compare the difference in average picking time and stockouts. To validate our measure(s) of *less-busy* store hours, we perform a field study in a grocery supermarket in Europe, under controlled settings. Results from the validation study corroborate our (treatment) measure of less-busy store hours. Next, we design an algorithm to schedule *non-urgent* orders to *less-busy* store hours, while imposing constraints to replicate the realities of our setting (for example, orders consisting of frozen and chilled products are picked up right before the delivery deadline, and not ahead of time). We perform a simulation to determine the productivity and service quality improvements resulting from the task scheduling of non-urgent orders to less-busy store hours.

Next, although scheduling of '*urgent*' orders, or orders containing frozen and chilled products might not be possible, it might be possible to reduce the shopping time by reducing (or eliminating) the queuing time. Two ways to do so could be using *self-checkout counters* (instead of regular checkout counters), and utilizing *scan-on-the-go* services using the store's mobile app. Therefore, for our second objective, we determine whether directing 'urgent' orders to stores that have either of these two features, could help improve order productivity, by significantly reducing shopping time. Although this approach is unlikely to have any effect on service quality, it could be utilized to achieve quicker turnarounds for urgent orders. Our final objective is to determine whether we could reduce the delivery time and delays of those orders that can neither be scheduled, nor be picked up from stores with self-checkouts or scan-on-the-go features. Precisely, we examine whether adopting a separate order picker and order deliverer model (which is commonly followed by quick commerce grocers), could reduce overall order (pickup and delivery) time. We utilize findings from an A/B test, conducted by our focal organization, to test this proposition. Specifically, by utilizing orders delivered in a single city in Europe, we match similar orders where a separate picker and delivery person were utilized (treated), to those orders where a single shopper performs both picking and delivery (control), and compare the average order delivery time and delays across both approaches.

We test our research questions using a comprehensive dataset from a European on-demand grocery platform and find that scheduling non-urgent orders to less-busy store hours can lead to improvements in picking time, without disrupting the quality of service delivered significantly. We find further evidence regarding the productivity benefits of utilizing mobile scan-on-the-go features and self-checkouts while picking urgent orders. Lastly, we report a positive impact on productivity (reduced delivery time and delays) by utilizing a separate picker and deliverer model for order pick-up and delivery. We conduct several robustness checks to reaffirm the veracity of our findings.

In conclusion, we contribute to a recent but growing literature on improving picking productivity of third-party personal shoppers, picking orders from retail stores. We highlight the benefits of avoiding customer encounters on both picking productivity and service quality. Lastly, we propose alternatives to practitioners (managers of ultra-fast grocers) to reduce additional investments in infrastructure and labor, and yet maintain desired levels of productivity and service quality.

2.3 Related Literature

In this paper, we examine the implications of reduced store traffic and fewer customer encounters; shorter queuing time; and reduced effort from task distribution (separate picker, deliverer) on picking productivity and service quality of third-party pickers. As such, our work is related to four separate fields of research: Order picking in retail stores and the impact on productivity due to picker types; The impact of product replenishment during slack periods on stockouts; The impact of queuing time on productivity; and the impact of task distribution on productivity and quality.

2.3.1 Order Picking in Retail Stores - Impact on picking productivity

Traditionally, order picking in retail has been examined in warehouse settings as opposed to picking in retail stores. In-store picking differs from warehouse picking in several aspects - the most prominent ones being the presence of other customers inside the store; and second, in-store picking requires the picker to travel to each product location (picker-to-parts) instead of the parts-to-picker mechanism that is common in modern warehouses

(Bozer and Aldarondo, 2018). Among existing studies on order picking from retail warehouses, Batt and Gallino (2019) is particularly relevant as the authors focus on picking of individual items for direct shipment to end users, which is similar to product picking (and delivery) in on-demand grocery. The study examines the retail order picking process and demonstrates that pick times increase by almost 16% as the search task becomes more difficult. In our setting, the search task becomes difficult due to the presence of (interference from) customers in the store, and a non-ideal picking environment which makes locating products more difficult. Furthermore, unlike warehouses that have well-designed storage areas for picking specific SKUs, most grocery stores do not have designated picking areas that could improve picking ease.

Next, we discuss the complexities of in-store order picking as a result of recent advancements in online retail such as the *the buy-online, pick-up-in-store* (BOPS) channel. In the BOPS method, customers purchase the products online, which is then picked-up by in-store employees. The customer later collects the products from the store, or a pre-specified location (Gallino and Moreno, 2014; Glaeser et al., 2019; Vyt et al., 2022). Another recent stream of literature has examined the picking performance of in-store pickers and suggested better picking (optimization) strategies (Chou et al., 2021a,b; Seghezzi et al., 2022). However, with the advent of on-demand grocery delivery, retail stores now consist of three different forms of pickers - customers, in-store employees, and third-party (platform) pickers, which in-turn could affect both picking efficiency, and the picking experience. To the best of our knowledge, Neves-Moreira and Amorim (2022) is the only work that has examined the in-store picker and customer interactions, and how that might affect picking efficiency. The authors propose a method to solve the dynamic in-store picker routing problem. However, research on whether such AI-enabled algorithms can benefit third-party pickers of online grocery platforms is scant. A recent study by Knight et al. (2023) has examined the implications of shoppers' ability to navigate stores using AI technology on order outcomes. The authors find that application of AI could have productivity benefits, but there are limits to the adoption and use of such technology, as overuse can even lead to lower productivity outcomes.

Our work relates to the emerging literature that examines how to improve order picking efficiency in retail stores. Specifically, we examine how picking in the presence of other customers could impact picking productivity, and whether scheduling of pickers to less busy store hours could help minimize some of the unwanted customer interference, while maintaining the service quality levels.

2.3.2 Stockouts, Replenishment, and Service Quality in *less-busy* store hours

Our work also relates to existing literature that examines the effects of retail stockouts and/or proposes solutions to mitigate the problem. In particular, we examine in-store factors, such as store replenishment, which could be responsible for stockouts. In this paper, we suggest alternatives (to third-party shoppers) to avoid customer encounters and improve picking productivity. One of the alternatives is scheduling of non-urgent orders to less-busy store hours. However, a concern regarding order picking in such hours is that several products (mostly fresh produce) are stocked-out during these hours. As a result, most large grocery supermarkets restock (or replenish) their items daily during these hours. While replenishment mostly happens at the end of each day, smaller items (and fresh produce) are often restocked within the day (Instacart, 2021) during less-busy periods. As such, it might be advisable for third-party order pickers to avoid such hours to not encounter product stockouts.

Among existing research that discuss the effect of stockouts and replenishment on picking performance, Gruen et al. (2002); Corsten and Gruen (2003) and ECREurope (2003) find that *store ordering causes* and *store replenishment causes* are the root *in-store* causes of out-of-stocks (OOS). The store replenishment-related causes of OOS imply that the product is in-stock within the store or the back room, but not on the store shelves (Aastrup and Kotzab, 2009). In most supermarkets, although the back-room replenishment might happen on-time, the replenishment of the store shelves is not so frequent, and often occurs once in a day (Instacart, 2021). Aastrup and Kotzab (2009) further mention that shelf replenishment is one of the major cause of in-store stockouts, contributing to 25% of the total number of stockouts, globally. In line with our study, there is a growing stream of literature that examines *phantom stockouts*, which occur when an item is physically available on the store shelves but cannot be found. Phantom stockouts are more prominent when picking is performed by third-party pickers, unknown to the store settings (layout). Ton and Raman (2010) examine stockouts in retail and find that phantom stockouts are caused when store navigation becomes challenging for pickers, or when in-store employees commit errors (or delay) while replenishing stock. In a recent study, Knight et al. (2023) examine a 'specific class of phantom stockouts' called '*accidental*' stockouts, which occurs when an item is physically present in the store shelves, but the shopper is unable to find it, either due to inexperience, or a complex store layout.

Our work adds a new direction to this stream of literature by examining the effect of less-busy store hours on stockouts. We conceptualize that the occurrence of improperly replenished shelves (and higher stockouts) is more probable during *less-busy* store hours. In-store employees often select these hours to replenish empty shelves, owing to lower on-shelf availability (higher number of stocked-out products). Our objective is to examine the magnitude of this effect for third-party (platform) shoppers and propose alternatives such as rescheduling of picking time to avoid stockouts and improve the overall service quality.

2.3.3 Impact of store traffic and queuing time on productivity

There is abundant literature on the impact of store traffic, and queuing on productivity and performance in retail stores. Perdikaki et al. (2012), for example, examine the relationship between store traffic and conversion rate. The authors find that higher store traffic does not translate to higher customer conversion. In fact, increases in traffic can lead to decreases in conversion rates, and this lower conversion rate is associated with a decrease in future traffic growth. In another study, Lu et al. (2013) analyze how waiting in queues of a retail store affects customers' purchasing behavior. The authors find that customers focus mostly on the length of the queue rather than the speed at which the queue is moving. The findings of this study are relevant in online grocery retail, where shoppers can batch pick multiple orders from the same store. Longer queues might lead the shoppers to abandon the queue and wait to pick other orders that they could batch. Personal shoppers may self-select into picking or queuing tasks, which may influence the picking time.

Our study fits into this stream of literature as in online grocery shopping, higher store traffic during busy hours leads to increased customer interference, which in turn increases the time required (by personal shoppers) to search for individual items. Past research has demonstrated that as the search task becomes more difficult, picking times increase (Batt and Gallino, 2019). Higher customer presence during busy hours may also increase waiting time in queues. Kc and Terwiesch (2012) demonstrate that servers (at counters) speed up when the system load increases. However, such temporal increases in speed cannot be sustained indefinitely, and hence servers slow down when the load remains high over an extended period of time (Kc and Terwiesch, 2009). During peak hours in a grocery store, as the number of customers in the store increases, the system load increases, and remains so for a considerably long duration. This could decrease the productivity of the servers, thereby increasing queuing time. The combined effect of increased search time, and longer queues, leads to an increase in the shopping time (which is the time required to pick and pay for items shopped). Therefore, shopper productivity will be higher, i.e., shopping time may reduce when orders are picked in a *less-busy* store (in *non-peak* hours) due to low customer interference, and shorter queuing time.

2.3.4 Impact of division of labor and task separation on productivity

Division of labor describes “the splitting up of a complex task into a number of specialized, simpler tasks” to increase efficiency of processes through specialization (Littek, 2001). Originally studied in the context of manufacturing and production, one of the main reasons for the separation (and specialization) of tasks was to achieve an improvement in productivity through the technical division of labor. An early work by Kilbridge and Wester (1966) examines the problem of the “economic extent of the division of labor”. Specifically, the authors evaluate the optimum cycle time that produces the lowest unit direct labor cost of assembly. Among studies that examine the counter-view of the impact of division of labor on productivity, Dobson et al. (2009) examine the division of labor, and profitability of primary care physician practices. The authors demonstrate that although task separation might be useful up to an extent, the division of labor introduces coordination costs. The authors assess the net effect of *task separation* and conclude that many physicians gain limited financial benefits from delegating work to their support staff. Their study suggests that smaller practices with fewer task segregation is better and more viable. In another paper, Belavina et al. (2017) investigate the division of labor when the tasks performed are novel and non-repetitive in nature, and argue that there should be a general tendency to cluster and allocate tasks by the intermediate objects they form (i.e., an object-based partitioning), rather than using the similarity of activities across different objects (i.e., an activity-based partitioning). However, when the task structure and the final product are both non-decomposable, only an *activity-based* task partitioning is possible. As such, activity-based division should be more common in the context of services. One of the more recent studies by Raveendran et al. (2022) examine the division of labor through *self-selection*. The authors examine the conditions under which self-selection-based division of labor can outperform the traditional allocation of work by managers. The results of the study demonstrate that self-selection is more beneficial compared to traditional staffing practices, where the employees are skilled in only a limited range of tasks, the task structure is decomposable, and the availability of the employee is uncertain. The results of this study are particularly relevant for gig-work-based online platforms where workers can self-select both tasks and work hours, and their actual availability is uncertain.

In the context of on-demand grocery picking, ultra-fast grocery platforms operate on the *separate picker and deliverer* model. This task separation (activity-based) helps save valuable time as the order is picked by an in-store employee, who works in the dark store, and is already packed and ready-to-go when the deliverer arrives to collect the order from the store. Furthermore, it minimizes worker fatigue as the personal shoppers now have to perform a single, specialized task. However, grocery picking from retail stores by third-party pickers is different from quick commerce, and therefore the benefits of task separation remains unexplored in this setting. Specifically, whether there exist any limitations beyond which task separation might be detrimental, requires attention.

2.4 Data and Empirical Strategy

2.4.1 Research Setting

Our research setting is an online grocery platform based out of Europe. The business model of the platform involves a customer placing an order using the mobile app and selecting a delivery slot. The order is assigned to a personal shopper (henceforth referred to as ‘*shopper*’) who is available to work in the specified hourly slot. The shopper then decides whether to accept the order. If the shopper accepts, she picks the order from the specified store, and delivers it to the customer.

2.4.2 Data Overview

Our dataset includes 2.58 million orders delivered in a period of 33 months, starting from January 2019 to September 2021. Each order consists of a set of unique identifiers for the order (*order_id*), the store (*store_id*), the customer (*user_id*), and the shopper (*courier_id*). Timestamps of order creation, first item picked, last item picked, payment at counter, delivery end, and delivery deadline are available for each order. Within each order, information is available for each of the individual items, which includes *item_id*, description of the item categories (e.g., for item_id: 101, category_1 description: milk; category_2 description: almond milk; category_3 description: 0 sugar almond milk), and the *brand_id*. Additionally, the company asks customers to provide a substitution preference for each item (1: do not substitute; 2: substitute after calling; 3: substitute without calling) while placing the order in case the item goes out-of-stock. We also have information regarding the status of the item (*Not found*: when an item is missing; *Purchased*: when the exact item requested is purchased; and *Replaced*: when an item is substituted) is also available. Finally, we have information regarding the price of each individual item, compensation received by the shopper for each order, information regarding the postal code (of the store and the customer), gender (of both shopper and customer), and age (of both customer and shopper). We utilize the individual datasets to create a consolidated dataset consisting of all the information at an order level.

The consolidated dataset has order-level information regarding the number of stockouts, substitutions, missing items, and items substituted when requested by the customer. Furthermore, a count of category-level items is also available. For example, our dataset has 22 *category_1* items such as meat, fish, dairy, frozen, vegetables, and so on. So we know how much meat, fish, dairy, and other products are present in the order. Such detailed information is useful to identify and match similar orders. We will discuss more in the model estimation section (4.1) of the paper.

2.4.3 Dependent Variables

Picking productivity. We measure productivity as the time required to pick an order (*pickingtime*). Picking time is the time elapsed since the shopper picked up the first item till the shopper picked the last item of the order, and does not include queuing time. We also repeat the analysis by using shopping time (*shoppingtime*), which is the sum of picking time and queuing time. For batched orders, the picking time overlaps between orders. Therefore, we calculate the picking time of a batched order as the difference between the last item picked timestamp of the last batched order, and the pick-start time of the first batched order, multiplied by the ratio of the number of items in the focal order and the total number of items in all orders combined. Specifically, let us assume there are 2 batched orders (X and Y). Order X contains 15 items, and Order Y contains 20 items. The pick-start time of order X is 9:15 a.m., and the pick-start time of order Y is 9:25 a.m., while the pick-end time of order X is 9:45 a.m. and the pick-end time of order B is 9:50 a.m. So, the picking time of order X will be (9:50 a.m. - 9:15 a.m.)*(15/(15+20)), i.e. 15 minutes (instead of 30), and for order Y it will be 20 minutes. As a robustness check, we also conduct this analysis by simply subtracting the pick end time and pick start time, without adjusting for the overlapping times.

Service Quality. We measure service quality as the number of stockouts in an order (*num_stockouts*). As a robustness check, we also used the number of missing items in an order (*num_missing*).

2.4.4 Treatment Effect

One of the objectives of this study is to determine whether order picking in *less-busy* store hours can help improve picking productivity (by reducing customer encounters) and service quality (by minimizing the occurrence of stockouts). We expect that during the day, when similar orders are picked from the treated (*less-busy*) store hours, the picking productivity and service quality should improve. However, obtaining a direct measure of a busy store is infeasible in our setting, as the third-party platform does not have access to actual store traffic data. As a result, we employ two different proxies (measures) for our main treatment, which is whether the store hours are *less-busy*/*empty* (1) or not (0). Our first measure (*emptystore_queue_store*) utilizes queuing time as a proxy of how busy the store is. If the queuing time of an order is below the 10^{th} percentile of the queuing time across all orders picked from that store, the order is said to be picked during a 'less busy' store hour.

However, it can be argued that a store can be less busy, and yet have a longer queuing time because only a few checkout counters are open. Since we do not have information regarding the number of open checkout counters, we employ a second proxy for *less-busy* store hours, which is based on busy vs non-busy picking hours (*emptystore_hour_store*). Each order in our dataset consists of a pick start hour. For e.g., if picking begins at 9:13 a.m., the pick start hour is 9. The hours during which at least 80% of the total daily picking activity occurs (begins) within a store, are denoted as *busy* store hours. These hours correspond to 7% (or higher) daily store traffic from third-party order pickers. The remaining hours are denoted as *less-busy* hours. Hence, we use this variable as the second measure to identify less busy store hours. Since it is possible that these hours differ by store, and store location within the city, we implement a store-specific measure of *emptystore_hour_store* (similar to the approach used in the queuing time-based measure).

However, we note that even this measure could be prone to errors as the *emptystore_hour_store* measure is based solely on the number of personal shoppers (third-party pickers) present in the stores in any given hour, and does not account for actual day, and hour-wise store traffic which includes both customers and in-store employees (in addition to personal shoppers). Therefore, to demonstrate the robustness of our results beyond a rough treatment metric, we employ a continuous treatment, wherein we model the degree of *busyness* in any given *store-hour* as an increase in queuing time. Specifically, we employ a generalized version of inverse probability weighting (IPW) for continuous treatment. We display the results of this estimation in the results section (5). However, recall that there is a concern regarding utilizing queuing time as a proxy of *store busy hours*. Therefore, we also validate the queuing time measure of *less-busy* store hours (*emptystore_queue_store*) by conducting a validation study, which is discussed in detail in section 6.

2.4.5 Matching Variables

We match orders based on order-item similarity (matching on 22 *Category-1* classifications), number of items, day of the week, and whether the order was batched (simultaneously picked with other orders) or not. The picking time of two similar orders might also differ based on the picker's (personal shopper's) prior experience, and same-day-experience (number of orders picked so far on the focal day), which accounts for both within-day learning and fatigue. Therefore, we include picker's prior experience and within-day experience in the matching estimation.

2.5 Econometric Analysis

In this section, we examine how *less-busy* store hours (treatment) affects the picking productivity (order picking time and shopping time) and service quality (number of stockouts) in an order (outcome). First, we describe the matching estimation employed in the analysis and its application to the empirical setting. Next, we discuss some robustness checks to ensure that the effects we report are robust to alternative explanations.

2.5.1 Matching Estimation

We are interested in measuring the effect of the treatment (order picking in a *less-busy* store) on an outcome, which in our case is the productivity and service quality of orders delivered. Specifically, we are interested in the average treatment effect (ATE), which is the mean difference between the outcomes of the treated and untreated (control) orders. The matching procedure creates pairs of *identical* (matched) orders with similar observable characteristics. The objective is to address the concern regarding the fact that the two populations- treated and control, are indeed different (Bell et al., 2020). Matching estimators are based on “the comparison of models that are similar” to each other, “with the exception that one receives the treatment” while the other does not (Moreno and Terwiesch, 2015). An assumption that is required while using this estimation methodology is that we observe sufficient covariates and any remaining influence on the treatment is uncorrelated with the outcomes. Later in the robustness checks, section 5.1, we discuss what could happen if the treatment was, in fact, correlated. We propose an endogenous treatment effect model to deal with such concerns. Our estimation is based on a propensity score matching (PSM), following Bell et al. (2020) and Moreno and Terwiesch (2015). Propensity score is the “estimated probability of receiving a treatment”. We compare the outcomes (picking time, shopping time, and number of stockouts) for orders picked from a *less-busy* store (treatment) and orders picked from a *busy* store (control) that have a similar propensity score. We estimate the probability of treatment using a logit model. Note that we also run a similar analysis using a probit model. We employ a 1:1 match and a caliper of 0.2 following Austin (2011). We include order-item similarity, number of items in the order, day of the week, same-day experience, and prior experience. After matching, we compare picking time, shopping time, and number of stockout for treated versus control store traffic as follows:

$$Pr(emptystore\_queue\_store = 1|Z) = \frac{1}{1+e^{-\beta Z}} - (1)$$

where, Z denotes the variables mentioned in section 3.5.

Later, in the robustness checks, we perform a nearest neighbor (NN) matching using the Mahalanobis distance of the covariates. We bias-adjust the continuous covariates for this estimation, following Abadie and Imbens (2006, 2011), who demonstrate that the NN matching estimators are inconsistent when matching on two or more continuous covariates and propose a consistent, bias-corrected estimator. We further perform estimations using the inverse-probability weighting (IPW) method (Hirano et al., 2003), regression adjustment (RA), and the inverse probability weighted regression adjustment (IPWRA) estimators, following Moreno and Terwiesch (2015)

2.6 Results

We present the findings of the first propensity score (PS) matching estimation - i.e., the effect of *less-busy* store hours on productivity and service quality.

Table 2.1: PSM - emptystore_queue_store

	(1) pickingtime	(2) num_stockout	(3) shoppingtime
emptystore_queue_store	-1.877***	0.030**	-9.783***
	(0.039)	(0.011)	(0.039)
N	2,579,112	2,579,112	2,579,112

$^{+}$ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 2.1 presents the results of the propensity score matching (PSM) estimation based on the "*queuing time measure*" of a busy store. As expected, we observe that both picking time and shopping time (i.e., sum of picking and queuing time) for orders shopped during *less-busy* store hours is lower, but the number of stockouts is slightly higher (i.e., service quality is lower). Specifically, we observe a 1.9 minute reduction in order picking time per order (significant at 1%), and a 9.8 minute reduction in overall shopping time per order (significant at 1%) when compared to a similar order shopped from a busy supermarket. However, the number of stockouts increases by 0.03 items (significant at 5%). A probable justification of this effect could be that while the stores have relatively less traffic in the *less-busy* hours, they also have higher stockouts due to non-replenishment of store shelves. In most stores across Europe, replenishment usually occurs once or at the most twice a day, usually in early morning or late evenings. During the day, if products are stocked-out, the chance of it getting replenished immediately, is low. However, stockouts might not occur in all such hours when store traffic is low. Therefore, while framing our scheduling algorithm, we specifically target those hourly slots, which are characterized by fewer stockouts.

Table 2.2 presents the results of the propensity score matching (PSM) estimation based on the "*less-busy hours measure*" of a busy store.

Table 2.2: PSM - emptystore_hour_store

	(1) pickingtime	(2) num_stockout	(3) shoppingtime
emptystore_hour_store	0.031	0.377***	-0.326***
	(0.033)	(0.008)	(0.038)
N	2,579,112	2,579,112	2,579,112

$^{+}$ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

We observe an insignificant change in the order picking time while utilizing this measure. The overall shopping time per order reduces by 0.33 minutes (significant at 1%) when compared to a similar order shopped from a busy supermarket. However, the number of missing items increases by 0.37 items (significant at 1%).

Despite the results being mostly supportive of our expectation, we realize that the *less-busy* hours measure might not be the most appropriate, owing to the fact that it does not account for actual store traffic, which includes customers and in-store pickers as well. Therefore, in the next section, we validate whether our first measure, based on *queuing time*, is in fact representative of store *busyness*. To establish the internal validity of our treatment variable - *emptystore_queue_store*, we conduct a validation study (section 6) in a single store of a popular store brand in Europe to observe the differences in productivity and service quality among *busy* and *less-busy* hours, while controlling for external sources of variance. The validation study further serves the purpose of answering the questions pertaining to whether utilizing self-checkouts or *scan-on-the-go* features could help reduce the queuing time, thereby reducing overall shopping time (of the order).

2.6.1 Robustness Checks

To test the sensitivity of our results, and to ascertain that our results are not reflective of a chosen methodology, we conduct several robustness checks. We perform matching estimations using the NNM (Appendix, Table B.1), IPW, RA, and IPWRA estimators (Appendix, Table B.2). Results of the analyses demonstrate full support (with comparable results) to our main analysis. Moreover, our findings using the original PS method were robust to changes in the model used (probit instead of logit), number of matches per treated unit, and the caliper - i.e., the maximum distance at which 2 observations are a potential match (Appendix, Table B.3).

Next, we use an *alternative measure* for the dependent variable, service quality. We use the number of missing items (*num_missing*), instead of the number of stockouts (Appendix, Table B.4). Results of the analysis are comparable and demonstrate full support to the main analysis. We also conduct a robustness check for the treatment variable, the *queuing time*-based measure of a *less-busy* store. In our main analysis, we assumed a queuing time less than the 10^{th} percentile could be denoted as a *less-busy* store. In this section, we use a measure based on the 25^{th} percentile of queuing time to observe whether the effects are valid under an alternative specification. The results demonstrate full support to the main analysis. Next, instead of a rough binary treatment measure of a *less-busy* store, we estimate the continuous treatment effect (an increase in queuing time) on the outcome variables. We utilize the generalized propensity score and dose response functions to estimate the effect. We include both linear and quadratic terms of our continuous treatment measure (*queuingtime*). Table 2.3 below presents the results of the estimation:

Table 2.3: GPS (Dose Response) - pickingtime vs continuous queuing time

	(1) pickingtime	(2) num_stockout	(3) shoppingtime
queuingtime	-2.061***	-0.340***	-1.061***
	(0.009)	(0.002)	(0.009)
queuingtime*queuingtime	0.046***	0.007***	0.046***
	(0.000)	(0.000)	(0.000)
pscore	-27.219***	-3.170***	-27.219***
	(0.244)	(0.058)	(0.244)
pscore*pscore	-10.582***	-2.510***	-10.582***
	(0.268)	(0.064)	(0.268)
queuingtime*pscore	5.537***	0.790***	5.537***
	(0.010)	(0.002)	(0.010)
constant	29.351***	4.102***	29.351***
	(0.056)	(0.013)	(0.056)
N	2,579,163	2,579,163	2,579,163

$^{+}$ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

The results demonstrate the presence of a curvilinear relation of queuing time (as a treatment) on the outcome variables: picking time, stockouts, and shopping time. Specifically, the coefficient of the squared term of queuing time demonstrates that when queuing time is longer (i.e., the store is busier), the order picking time and shopping time will be longer, and the number of stockouts will be higher, which is in line with our main estimation.

Lastly, we address the concern that our treatment could be endogenous. This concern could arise because there could be unobserved factors, such as experienced gig workers self-selecting the hours with lesser store traffic, which could impact both the treatment and the outcome. Therefore, in this section, we observe the endogenous treatment effect using an instrument, *avg_queuetime_perorder*, which is the average queuing time per order (before the focal order) in a given store. The average queuing time per order is representative of the store traffic, and therefore the higher the average queuing time in a store, the fewer the chances are of an empty store. However, the average queuing time per order in a store is unlikely to directly impact the picking time and number of stockouts in the focal order. Therefore, our instrument satisfies both the relevance criterion and the exclusion restriction. We present the results of the estimation below.

The results from the estimation further validate the original estimations. The wald-test of endogeneity is significant, suggesting that our regressor (treatment) is endogenous. We observe that *less-busy* store hours (*emptystore_queue_store*) reduces order picking time, and shopping time, while increasing the number of stockouts.

2.7 Validation Study

2.7.1 Study Design and Objectives

The objective of the validation study is to establish the accuracy of our *treatment* measure. Typically, in a validation study, the accuracy of a measure is compared with a gold standard measure (Fox et al., 2020). In this case, the *gold standard* would have been actual store traffic data (footfall count, and/or video cameras at checkout counters). Owing to the lack of such measures, we employ a field study in a controlled environment, similar to a lab experiment. We caution the readers at this point that this study is not a field or lab experiment, and hence does not include manipulation checks or pre-registering

Table 2.4: Endogenous treatment effect

	(1) pickingtime	(2) num_stockout	(3) shoppingtime
num_item	0.553***	0.099***	0.605***
	(0.000)	(0.000)	(0.001)
day_of_week	0.014***	-0.104***	0.062***
	(0.004)	(0.001)	(0.005)
batch_pick	7.408***	-0.139***	10.933***
	(0.029)	(0.007)	(0.034)
ln_experience	-2.349***	-0.269***	-2.593***
	(0.005)	(0.001)	(0.006)
same_day_experience	-0.586***	0.106***	-0.600***
	(0.003)	(0.001)	(0.004)
1.emptystore_queue_store	-8.023***	5.055***	-20.397***
	(0.083)	(0.006)	(0.074)
constant	25.811***	1.371***	37.669***
	(0.036)	(0.008)	(0.041)
emptystore_queue_store			
avg_queuetime_perorder	-0.038***	-0.014***	-0.059***
	(0.001)	(0.000)	(0.001)
constant	-0.762***	-0.945***	-0.521***
	(0.006)	(0.005)	(0.006)
athrho	0.251***	-1.171***	0.408***
	(0.004)	(0.002)	(0.003)
lnsigma	2.515***	1.240***	2.680***
	(0.001)	(0.001)	(0.001)
Observations	2,579,112	2,579,112	2,579,112

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

the study. Our objective is simply to validate whether a longer (shorter) queuing time is representative of a busy (less-busy) store.

Next, we discuss the study design and the objectives behind it. First, we ensure that the basket of items closely resembles an actual order basket. We list 20 items which are the same across all orders, and cover a range of products such as fresh fruits and vegetables, meats, dairy, laundry detergents, rice, bread, spices, beverages, snack foods, and skin care (see figure B.1 of the Online Appendix). The variety of items resembles an actual order basket and requires our participants to cover a broad array of shopping aisles, which ensures that the order picking times would be similar to those in our actual dataset. To ensure representativeness of the study results, we selected a store brand that has presence across Europe. Our chosen store has both normal checkout, and self-checkout counters. Additionally, the store's mobile app allows customers to scan items on-the-go.

To conduct the study, we hired two participants (not familiar to each other, and did not have prior experience of working as a third-party order picker) who were assigned the task to pick up a list of groceries from a single supermarket at three different hourly slots within the day. The study was conducted across all days of the week, for a period of four weeks. Prior to the first week, the participants were prepped regarding the task, and were accompanied by one of the authors to ensure that the participants understood the task and collected the required information. The data during this one-week trial phase was not utilized. Furthermore, the trial ensured that the participants did not start performing better at a later stage as they gained familiarity to the supermarket. The actual data was collected between 16^{th} June 2022 to 13^{th} July 2022. The activity was conducted for four consecutive weeks, across all days of the week, to capture variance across days and weeks. The participants were compensated €5 for each instance of shopping. Overall, each participant earned €420 in 28 days. The total allocated budget for both participants was €840. Since the study involved human subjects, the authors pre-approved the study from the Institutional Review Board (See figure B.2 for the participant information sheet).

Each participant was assigned three primary tasks - First, they were required to pick (scan using the mobile app) 20 different items provided in a list, and record the pick start time, the pick end time, the number of items found, and the number of items substituted. Participants also had to record a perceived score for *store busyness* on a scale of 1-5; 1 being the least busy. To capture the service quality measure, brands names were provided in the item list. In case an item was missing, the participants recorded the number of items for which they picked (scanned) a substitute product. To ensure that the availability of the brand was not influenced by external factors, we specified popular brands which could easily be found in our selected store, and were less likely to go out-of-stock. In the next step, the participants had to stand in the checkout queue and record the queuing time, and the number of open checkout counters. Finally, the participant had to stand in the self-checkout queue and record the queuing time, and the number of open self-checkout counters. We further clarify that the participants did not actually pick (or pay for) the items in the list. They simply scanned the items, recorded the timestamps, and then stood in the queue. The participants were instructed to quit the queue as soon as the person in front of them started paying. While using the self-checkouts, if the participants observed unoccupied self-checkout machines, they recorded the queuing time as zero; else, they recorded the queuing time similarly (as described before).

Here, we discuss the purpose of the study design. First, to avoid researcher bias, our study required hiring an external candidate who had no prior information regarding the objectives (and end-goals) of the study. We did not hire multiple participants owing to budget constraints and the availability of viable candidates. Furthermore, a single candidate helps eliminate sources of variability across participants. However, to ensure inter-participant agreement, instead of a single candidate, we hired two participants. Orders were picked on three different slots (9 a.m.-1 p.m.; 1 p.m.-5 p.m.; and 5 p.m.- 9 p.m.) within the day. The idea of selecting three different time slots was to compare *busy* vs *less-busy* store traffic during the morning, afternoon, and evening work shifts. 9 a.m.-

1 p.m. on weekdays corresponds to typical work hours; 2 p.m.-5 p.m is often the time referred to as '*siesta*' or a break period, when most people take a break for lunch, or close their businesses for a brief rest period; and finally 5 p.m.-9 p.m. is the time when people return from work and are likely to visit the stores. To ensure that we capture not just the differences across hours, but also across days, our participant performs picking on all seven days of the week. Finally, since the shopping behavior of customers could change based on whether it is the beginning of the month (which is when they receive their salary), or the end of the month, we collect data across four weeks to capture the variances across the month. An additional advantage of conducting the study across an entire month is that almost all hourly slots between 9 a.m.-9 p.m. get some visibility in our collected data.

Finally, we discuss the main objective(s) of the validation study. The primary objective of the study was to analyze whether the participant's perception of 'busy store' corresponds to a higher queuing time, while controlling for the number of open checkout counters. Specifically, we expect that if a participant perceives a store to be busy (less busy), the queuing time should be higher (lower), while controlling for the number of open checkout counters. This, in turn, would help us validate the "queuing time-based measure" of a "less-busy" store. Next, results from the validation study, combined with secondary data from the on-demand platform, could help us better identify the *less-busy* hourly slots. Finally, data from the queuing time estimates of regular checkouts vs self-checkouts provide evidence of the benefits of using self-checkouts.

2.7.2 Study Results

In this section, we present the results of the validation study. In table 5 below, we present the regression results of queuing time (dependent variable) vs workers' perceived measure of a *busy store*. We remind the readers that the participants of the study rate "store busyness" on a scale of 1-5; 1 being the least busy. We expect that the queuing time of a perceived *less-busy* store will be less. Specifically, a longer queuing time is representative of a *busy* store.

Table 2.5: Queuing time (regular checkout) vs Perceived store busyness

	(1)	(2)	(3)	(4)	(5)
busy_store_1to5	1.256***	1.335***	1.290***	1.286***	0.618***
	(0.102)	(0.160)	(0.085)	(0.085)	(0.145)
num_open_checkouts	-0.099	-0.184***	-0.153***	-0.153***	-0.090*
	(0.066)	(0.066)	(0.044)	(0.045)	(0.047)
constant	3.560*	5.720**	4.999***	4.983***	4.754***
	(2.003)	(2.302)	(1.429)	(1.433)	(1.437)
Observations	84	84	168	168	168
R2	0.688	0.673	0.681	0.681	0.732
AdjR2	0.680	0.665	0.677	0.675	0.724
Participant fixed effects				Yes	Yes
Slot fixed effect					Yes

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Models 1 & 2 includes 84 observations each (3 observations per day for 28 days) by participants 1 & 2, respectively. Models 3 to 5 include all 168 observations by both participants. Model 4 includes participant fixed effects, and Model 5 includes both participant and hourly slot fixed effects. Across all models, we observe that participants' perceived store busyness corresponds to higher queuing time (i.e., higher the busyness, longer the queuing time). We further observe that queuing time decreases if more checkout counters are open. Results from the validation study demonstrate that the *queuing time-based* measure of a *less-busy* store is accurate, and can be utilized to demonstrate the benefits of order picking in a *less-busy* store.

Next, to validate our original matching estimation (presented in Table 1), we examine the effect of the participants' perceived store busyness (*busy_store_1to5*) on order picking time. Table 6 below, presents the results of the nearest neighbor matching estimation. Models 1 & 2 include 84 observations each, by participants 1 & 2, respectively. Models 3 & 4 include all 168 observations by both participants. Model 4 includes both participant fixed effects. Furthermore, all models include hourly slot fixed effects. Across all models, we observe that participants' perceived store busyness corresponds to higher picking time (i.e., the busier the store, the longer the picking time).

Table 2.6: Picking time vs Perceived store busyness

	(1) picking_time	(2) picking_time	(3) picking_time	(4) picking_time
busy_store_1to5	1.869**	1.374	1.123*	1.768***
	(0.848)	(0.885)	(0.600)	(0.579)
constant	15.547***	15.159***	16.687***	16.212***
	(2.251)	(2.832)	(1.766)	(1.662)
Observations	84	84	168	168
R2	0.121	0.037	0.038	0.156
AdjR2	0.088	0.001	0.021	0.136
Participant fixed effects	No	No	No	Yes
Slot fixed effect	Yes	Yes	Yes	Yes

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Our second objective is to determine whether queuing time using self-checkouts is better (shorter) than queuing time in regular checkout counters. We perform a paired t test to observe the difference in queuing time means (self-checkout vs regular checkout). Table 7 presents the results:

Table 2.7: ttest for mean difference in queuing time (self-checkout vs regular checkout)

	actual	predicted	ATE
mean	0.613***	5.667***	-5.054***
	(0.0796)	(0.130)	(0.152)
N	168	168	336

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

The results in column 3 of Table 7 demonstrate that the mean difference of queuing time in self-checkouts is significantly smaller than that of regular checkouts. Although, owing to the small sample size, and the fact that our data comes from a single store, we caution the readers towards the extrapolation (or generalizability) of our results. For example, in a recent blog, Allon (2021) caution customers towards the demerits of self-checkouts, especially in busy stores. He argues that the issue with self-checkouts is that its users are not well “trained” in terms of speed or product choices. For example, variable weight items or age-restricted items could increase order processing time. This, combined with higher congestion, could increase queue length. Future research using actual queuing time and store traffic data from multiple stores is required to comment about the benefits of using self-checkouts.

2.8 Order Scheduling Algorithm

In the previous sections, we discussed the benefits of order picking from a *less-busy* store in terms of improved picking productivity and service quality. We believe that the benefits of avoiding customer encounters and longer queues, proposed in this paper, are not only statistically significant, but also contain managerial relevance. None of the existing studies have examined the extent to which order picking time and stockouts could be reduced in the absence of dark stores or routing algorithms, by scheduling non-urgent tasks to less-busy store hours. Investigating this effect is especially important for online grocery platforms that perform picking from local supermarkets, and therefore often do not have access to product location or real-time inventory information.

In this section, we propose to minimize customer interference by scheduling non-urgent tasks to *less-busy* hours, which correspond to lower store traffic. In doing so, we use multiple constraints to reflect the realities of our setting closely. First, since all *urgent* orders need to be delivered within 1–2 hours, we do not schedule those orders. Therefore, the benefit from the proposed task scheduling algorithm is especially relevant for *non-urgent* or *scheduled* orders. Next, several grocery orders have ambient, chilled, and frozen items. If chilled and frozen items are picked up early, there could be quality issues due to improper storage conditions. Therefore, all orders that contain items whose quality or structural integrity might be affected by early picking (for e.g., meat, frozen dairy etc.), need to be treated as urgent orders and cannot be scheduled to an earlier hour.

Of the remaining orders, some are already picked in a less-busy hour, and therefore do not need to be scheduled. We therefore only schedule orders that are non-urgent, do not consist of chilled or frozen items, and whose picking is originally scheduled in a busy hourly slot. In the next step, we utilize our dataset to determine hours during which a store is *less-busy*, but not prone to stockouts. To do so, we select hourly slots characterized by lower queuing time (*emptystore_queue_store*) for each individual store. We do not utilize the *emptystore_hour_store* measure as it is not a true representation of actual store traffic. Moreover, according to the *emptystore_hour_store* measure, for most stores, *less-busy* hours coincide with late evening or mid-afternoon hours, which might also correspond to higher stockouts. On the other hand, the queuing time-based measure (*emptystore_queue_store*) is indicative of lower store traffic, but does not necessarily translate to higher stockouts, as such hours vary across individual stores and days in a week.

In the next step, for each *store_id*, the algorithm checks the ratio of *emptystore_queue_store==1* and *order_id* for each *pick_start_hour* ($\frac{emptystore\_queue\_store==1}{order\_id}$). Specifically, the algorithm calculates the average percentage of times a store was *less-busy*, in each pick-start hour. For e.g., for store_id 4507, the pick_start_hours of 7 a.m. and 8 p.m. were deemed to be the least busy, with 25% and 19% of *less-busy* hours respectively. This indicates that as long as an order is non-urgent, does not consist of frozen or chilled products, and has a delivery deadline after 8 a.m., we can schedule the picking to start at the 7 a.m. slot for store_id 4507. Note that while scheduling orders, we assume that there are a sufficient number of workers present in each slot to deliver the orders scheduled, and do not include constraints such as worker availability, or worker-specific characteristics (such as experience and fatigue) that could also impact order picking and/or shopping time. We include a provision to dynamically calculate the number of workers allocated to each hourly slot, within each store. This is particularly useful if an organization has a cap on the number of workers available to work in each hourly slot. The algorithm can use the max cap to allocate orders to the next *less-busy* slot (if the least busy slot is filled up).

We use the following linear equation (OLS) to predict the picking time, shopping time, and number of stockouts, and compare the predictions of the original dataset with that of the new schedule developed by the algorithm:

$$Performance_{is} = \alpha_0 + \beta_1 emptystore\_queue\_store_{is} + X_{is} + \varepsilon_{is} - (2)$$

where $Performance_{is}$ denotes one of the three performance measures for picking productivity (picking time, shopping time), and service quality (number of stockouts) of an order *i* picked from store *s*, $emptystore\_queue\_store_{is}$ denotes the queuing time-based measure of the treatment, which indicates whether a store is busy or not, and X_{is} denotes the set of controls, which is the same as the matching variables used in section 3.5 (and equation 1), but excluding the worker's prior and same-day experience. Table 8 below presents the predicted improvements in picking time, shopping time, and stockout reduction, by scheduling non-urgent tasks using our scheduling algorithm.

Table 2.8: Task Scheduling Algorithm - Comparison with baseline values

	(1) Mean
Percentage of change in *picking time*	-3.76%
Percentage of change in *shopping time*	-12.65%
Percentage of change in *missing items*	-1.05%

The results of the predictions demonstrate a 3.76% reduction in order picking time, a 12.65% reduction in shopping time, and a 1.05% reduction in missing items, by scheduling non-urgent (scheduled) orders. Overall, we find a positive managerial impact of task scheduling of non-urgent (scheduled) orders on productivity (picking and shopping time) and service quality (missing items).

2.9 Separate picker and deliverer

In the previous sections, we examined the impact of customer interference and queuing time on order picking time, shopping time, and missing (stocked-out) items. Next, we examined the benefits of order picking from a *less-busy* store and proposed a task (order) scheduling algorithm that could help reduce picking time and stockouts in scheduled (non-urgent) orders. Furthermore, we suggested the use of self-checkout counters and *scan-on-the-go* features to reduce (or eliminate) queuing time, thereby reducing overall shopping time. However, as mentioned before, our suggestions have restrictions. The scheduling algorithm is not applicable to urgent orders. As a result, the only way to improve productivity is by utilizing self-checkouts or mobile scan-on-the-go. What exacerbates this issue is the lack of self-checkouts or dedicated apps (that allow scanning and paying for items on-the-go) in several stores. This forces on-demand grocery retailers to invest in *dark stores* or in the development of expensive picker routing algorithms in collaboration with grocery stores. However, an often unexplored, yet effective technique of improving order productivity and service quality is to focus on the delivery (rather than the picking) aspect. Guha and Corsten (2023) demonstrate that gig workers, who pick and deliver orders from grocery stores, often get delayed while working longer hours. Specifically, the authors demonstrate that as workers' within-day experience through orders delivered (or hours worked) increases, the order delay time increases. One possible cause for this finding could be fatigue from performing different tasks for a long period of time. In section 2.4, we discuss how the division of labor and task separation could help improve productivity. Drawing from past findings, we expect that utilizing separate workers for order picking, and order delivery could not only help reduce worker fatigue, but also save valuable time in both activities. Furthermore, all on-demand grocers operating in a *"dark store"* model, utilize the separate picker & deliverer model, which allows them to be more efficient and achieve quick turnaround times. It would therefore be interesting to observe whether a similar effect can be obtained from in-store picking of on-demand grocery orders.

In this section, we investigate the impact of division of labor (by assigning separate workers for order picking and delivery) on productivity and service quality. We measure productivity using *deliverytime*, and service quality as using the *delay_time* (in minutes). A delay of zero (0) minutes signifies an on-time delivery. We examine our research question by utilizing an A/B test conducted by the online grocery platform which provided us the dataset (henceforth referred to as “company”). The company performed an A/B test in some select stores in one of the cities (henceforth referred to as “city A”) in which it operates, to determine whether a *separate picker & deliverer* model yields productivity benefits. In the data subset (for the selected stores in city A), there are 55,417 observations in the period during which the A/B test was conducted, and not all orders were subjected to this separation model - i.e., some orders had a single worker performing both picking and delivery tasks. This allows us to match nearly identical orders, which received (or did not receive) the treatment (*separate_picker_deliverer*), and identify the benefits of utilizing the separate picker & deliverer model. Specifically, orders in which a separate picker & deliverer model was utilized were assigned the value of 1 (i.e., received treatment), and those which did not, were assigned 0. Our main estimation is based on propensity score matching (PSM) using covariates such as whether the delivery distance (proxied using store’s postal code and customer’s postal code), day of the week, the delivery-end hour, and whether the order was batched or not. Results of the estimation are demonstrated in Table 9 below:

The results demonstrate that both the delivery time, and the order delay reduces if the activities are performed by separate pickers & deliverers. Since our estimation is based on an A/B test conducted in a single city and platform, we cannot comment on the generalizability of our results to other platforms and settings. Still, we believe that future research can build upon our work to explore the benefits of picker-deliverer separation on

Table 2.9: Propensity Score Matching - Separate Picker & Deliverer

	(1) deliverytime	(2) delay
ATE		
diff_pick_deliv	-2.163***	-2.832***
	(0.182)	(0.178)
Observations	55,417	55,085

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

picking and delivery performance.

2.10 Discussion and Conclusion

2.10.1 Discussion

Our work contributes to a growing stream of literature on improving the performance and profitability of online grocery retailers (Allon et al., 2023; Belavina et al., 2017; Dai et al., 2022; Guha and Corsten, 2023). In this study, we examine how online grocery platforms could improve order picking efficiency from retail stores. In particular, we examine whether scheduling tasks based on store traffic (*less-busy* vs *busy* store hours), while considering order-specific characteristics (such as urgent orders, and orders containing frozen and chilled items) could improve order picking time and reduce the chances of stockouts. We examine the tradeoff that although picking time will be shorter (higher productivity) in a less-busy store, less busy periods are often characterized by empty store shelves, owing to ongoing replenishment activities, which might lead to a poor quality of service, in terms of higher number of stockouts and missing items. We find evidence that while picking time reduces, stockouts might increase slightly. We then devise an optimal task scheduling policy by scheduling tasks in those hours that are characterized by lower store traffic. Specifically, we identify *less-busy* hours for individual stores by utilizing the average queuing time for each hourly slot. We demonstrate the benefits of utilizing our scheduling algorithm in terms of picking productivity and service quality. Furthermore, for orders where scheduling is not possible, we examine how utilizing self-checkouts and mobile *scan-on-the-go* can improve productivity by reducing the queuing time of ‘urgent’ orders. Lastly, for orders that can neither be scheduled, nor be picked in a store that has self-checkouts, we suggest employing a separate picker and deliverer model, where each individual is responsible for a single task (picking or delivery). Utilizing results from an A/B test, and performing a matching estimation, we demonstrate the benefits of the separate picker and deliverer mechanism.

2.10.2 Contributions and Managerial Implications

Our objective is to advance research in the field of online grocery shopping by third-party pickers from retail stores, by focusing on the detrimental effects of customer interference and queuing time on productivity and service quality, and suggest techniques to improve picking productivity and service quality. We quantify the benefits of picking during *less-busy* hours and make the following suggestions: First, we suggest online grocers to schedule picking of non-urgent orders during hours that coincide with less queuing time, rather than mid-afternoon or late evening hours, to avoid orders being more prone stock-outs. Second, we suggest online grocery platforms to schedule picking from those stores which are known to have self-checkouts and/or *scan-on-the-go* mobile apps. We also suggest training personal shoppers on how to use self-checkout counters, and ‘scan-on-the-go’ features to avoid queuing time. Finally, we suggest such platforms to enable a separate picker and deliverer mechanism, especially during busy store hours. This could help platforms achieve quicker turnaround times with fewer picking errors.

To put things in perspective, the matching estimations demonstrate a reduction in picking time (excluding queuing time) by 1.7 to 2 minutes. The average picking time of an order (in our dataset) is 26 minutes (for an order consisting of 30 items, and costing 78 euros, on average). The average number of orders delivered per day across all workers is 4 orders. However, for workers in the top 25% of work hours, i.e., those that behave as near full-time workers, the average orders per day per worker is 10 orders. Assuming a worker delivers 10 orders in a day, she can save 20 minutes or more, by picking orders during a *less-busy* store-hour, which is almost enough time to pick an extra order. To add to this, the worker could save an additional 11 minutes of queuing time (on average), by utilizing self-checkouts or *scan-on-the-go*. Furthermore, workers could reduce 2.2 minutes on delivery time, by utilizing a separate picker and deliverer model. The overall time saved by adopting these approaches could be sufficient to pick and deliver 1 to 2 additional orders per worker, per day, depending on the number of hours worked by the worker. Assuming at least 25% of all active workers in our dataset (1000 workers) are working near full-time in a day, and delivering 1 (one) additional order costing 78 euros, it translates to 78,000 euros of additional earnings per day (assuming that there is enough demand). Since the operating model does not involve owning dark stores, the additional earnings, and the profits that come from it, could be utilized in growing the online grocery business, and paying better wages to the gig workers.

2.10.3 Limitations and Scope for Future Research

As with all studies, our results are also subject to certain limitations. As stated before, our measures of busy store are prone to errors. While we have tried to address it to the best of our abilities by utilizing a continuous treatment measure, and by conducting a validation study to validate our measure of a *less-busy* store, we lose external validity in exchange for internal validity, when the study is conducted on a small sample. Future work can employ real-time store traffic data to examine similar questions. Future work in this field could also examine the ‘value’ of shopping in a ‘dark store’ and whether improvements in productivity, and service quality can mitigate the store maintenance expenses. We also urge authors to observe the effects on service quality more closely, for example, by employing different measures of service quality (picking accuracy), especially in case of substituted products, such as the closeness of substitutions to the original product requested. Finally, future work could also design simulations to examine how to incorporate worker availability in the scheduling algorithm to best allocate picking time slots for scheduled deliveries to personal shoppers in a way that optimizes both service quality, and productivity.

Chapter 3

The Impact of Service Quality and Customer Experience in Online Store Choice and Platform-Exit Behavior

3.1 Abstract

In recent years, grocery shopping using online platforms, which act as third-party aggregators for grocery store chains, has picked up pace. Unlike traditional retailers, online aggregators allow customers to select their store-brand of choice, without incurring significant travel and search costs. While this ease of switching allows customers to patronise multiple store-brands, they remain loyal to a few preferred store-brands, which comprise a majority of their purchases. In this study, we examine the factors affecting customers' intrinsic utility towards their 'primary' online store-brand, and its second order effect on primary store switching. In particular, we focus on customer-specific drivers, and online service quality drivers. Customers' shopping behaviors evolve as they gain online shopping experience. This affects the intrinsic utility towards their primary store brands, when service quality issues such as stockouts and substitutions occur. Our objective, therefore, is to examine the effect of online experience and service quality on the decision to switch stores, and to exit the platform, respectively. To investigate our research questions, we utilize purchase data from an online grocery platform, over a period of 33 months. We employ a Bayesian estimation of posterior probabilities to investigate customers' store-switching and platform-exit behavior; and a multinomial logit choice model to investigate factors affecting online store choice.

Our findings reveal that service quality parameters such as stockouts and substitutions play a significant role in customers' decision to switch their 'primary' online store-brand, or exit the platform's service. Interestingly, we observe that customers' online shopping experience has opposing effects on store-switching, and platform-exit, respectively. Specifically, when experienced customers encounter stockouts (or stockout-based substitutions), they are less-likely to switch their preferred store-brand. However, experienced customers are more likely to exit the platform after experiencing stockouts in the focal period. This finding indicates that experienced customers exhibit inertia towards their preferred store-brand, but are less-tolerant towards the online platform in case of service quality issues. Lastly, we also find that the probability to exit the platform's service is higher when a customer has had a recent 'primary' store switch. We discuss the implications of our findings for both retail stores, and the online platforms.

3.2 Introduction

With the growing popularity of online grocery platforms, there has been a shift in customers' perception of and attitude towards grocery shopping. Online grocery platforms act as aggregators of retail store chains, and provide customers with an online one-stop solution for all grocery-related purchases. Customers can use the platform to select the store brand of their choice, and place their order request. Alternatively, customers can search for a specific product and check for available alternatives across the different store brands. This ease of switching between store-brands has significantly reduced customers' search and travel costs compared to traditional grocery shopping from retail stores, or online grocery shopping using the website of specific store chains. Grocery retailers enrolled with the online platform often respond to customers patronising multiple stores, by offering them promotions and discounts, thereby trying to acquire new customers.

Past research has demonstrated that price promotions are often ineffective in switching the store to which a customer is primarily affiliated (Rhee and Bell, 2002). However, much of this effect is driven from the fact that customers are unwilling to "let go" of their prior store-specific knowledge of assortment and product locations. Therefore, whether customers would have an online 'primary' store is debatable, because factors such as distance, or knowledge of product assortment are not applicable in online platform-based grocery shopping. Melis et al. (2015) put forth that multi-channel shoppers, when shopping online, are inclined to select the online store belonging to the same store brand as their preferred offline store. It could therefore be expected that inexperienced (or first-time) users of the online platform-based grocery shopping would prefer to select an online store based on their prior knowledge of shopping in an offline store belonging to the same chain, and then modify their choice as they gather more online shopping experience, by incorporating their online store-specific utility, and the utility derived from using the platform. An important aspect, that has eluded prior research is the extent to which online service quality parameters, such as delays, stockouts, and stockout-based substitutions, can affect customers' utility while shopping groceries online, and whether such (dis)utility could affect a store-switch, or a platform-exit behavior.

Despite the growing body of research on switching behavior in online grocery shopping, an aspect that has been understudied is the effect of *push* factors such as service quality issues (delays, and stockouts) in past orders (Boyer and Hult, 2005), on customers' decisions - specifically, the decisions to switch to a different alternative within the same-service, or to exit the particular service. Prior research has demonstrated that the adverse impact of a stockout extends beyond the focal order to future purchases as well (Anderson et al., 2006). Therefore, while it is intuitive to assume that a recent stockout could cause an immediate online store-switch, whether this effect persists over the long-term, i.e, whether stockouts in the focal period could cause a switch in the 'primary' store of affiliation (in the next period), is less straightforward. The complexity of this problem further increases when we incorporate customers' online shopping experience and how grocery shoppers behave while shopping online (Anesbury et al., 2016). As customers become more experienced, they adjust the utilities derived from shopping in specific store brands, and hence their preferences towards the available store brands (Melis et al., 2015). Furthermore, it is likely that as customers' online grocery shopping experience increases, some customers develop a *stickiness* towards certain store brands and would therefore be more resistant to change their store of choice, while others might develop a *stickiness* towards the online platform itself, and would therefore be willing to switch stores, when faced with a service quality issue. Therefore, in this study we investigate the effect of stockouts on store-switching and platform-exit behavior, separately, when the online shopping experience of customers increases. Lastly, is it possible that customers might react differently if the stockouts were to occur in a product category where they spend less, as opposed to a stockout in a high-spend category. To accurately capture this effect, both measures of missing items and substitutions are weighted using each customer's category spend.

Finally, our study also examines the price aspect of online grocery shopping - i.e., the effect of product prices on switching or exit probabilities. Price of items (or price of the order basket) is a commonly used parameter while examining store-switching (choice) behavior (Rhee and Bell, 2002; Van Lin and Gijsbrechts, 2014; Melis et al., 2015). However, the effect of the 'value' associated with an online purchase is often overlooked. Different customers associate different values to product categories. For example, in a given time-period, the percentage allocation of amounts to be spent in each category would be different for each customer. Therefore, the ‘value’ of purchase in a given time period would be the sum of the amounts spent in each category, weighted using the customer’s spend ratio in those categories. We expect that while higher item prices in a store-brand could affect a store-switch, higher purchase values could have the opposite effect, following prior research that has demonstrated that past customer spending across categories is a positive signal of future customer spending (Anderson et al., 2024).

Our main objectives, therefore, are threefold: We first investigate the primary drivers of online store switching and platform-exit. Specifically, we focus on the main effects of observed service quality parameters - delays, stockouts (missing items), and substitutions; and the effect of the value of purchase made. Next, we investigate the effect of customers’ online shopping experience on store-switching, and platform-exit, respectively, when they encounter stockouts and substitutions in the focal time period. Lastly, we examine whether the factors that lead to a store-switch could be utilized to examine customers’ choice of the store brand (i.e., store choice). We examine our research questions using purchase data from an online grocery platform in Europe, over a period of 33 months. We employ a Bayesian estimation based on a set of priors (our parameters of interest) to investigate the posterior probabilities of a store-switch and/or a platform-exit. To investigate factors affecting customers’ online store choice, we employ a MNL choice model.

Our findings reveal that service quality parameters such as stockouts and substitutions play a significant role in customers’ decision to switch their ‘primary’ online store-brand, or exit the platform’s service. Interestingly, we observe that customers’ online shopping experience has opposing effects on store-switching, and platform-exit, respectively. Specifically, when experienced customers encounter stockouts (or stockout-based substitutions), they are less-likely to switch their preferred store-brand. However, experienced customers are more likely to exit the platform after experiencing stockouts in the focal period. This finding indicates that experienced customers exhibit inertia towards their preferred store-brand, but are less-tolerant towards the online platform in case of service quality issues.

Lastly, we also find that the probability to exit the platform's service is higher when a customer has had a recent 'primary' store switch.

This study contributes to existing literature on online store choice and store switching. To the best of our knowledge, we are the first to investigate (i) the drivers of platform (service)-exit when customers encounter service quality issues while utilizing an online platform to purchase their groceries, (ii) the effect of online purchase experience on 'primary' store switching, after encountering service quality issues, (iii) the effect of the 'value of purchase' on primary store switching and store choice, and (iv) the second-order effect of a recent primary store-switch on service-exit. From a managerial point of view, our results provide useful insights for both online aggregator platforms and retail stores enrolled with online platforms. In particular, we suggest retail stores (enrolled with online platforms) to reduce price promotions and discounts, and instead focus on providing suitable alternatives (substitutions) to products that are likely to be stocked-out, to retain their 'loyal' customers. We further suggest the online grocery platforms to adopt techniques through which picking efficiency (i.e, picking time and quality of picking) could be improved. Improved picking could lead to lesser delays, and higher substitutions in case of stockouts, which in turn, would help retain customers. Online grocery aggregators are facing challenges to retain their customers (Keaveney and Parthasarathy, 2001; Singh and Rosengren, 2020). Our suggestions provide direction to managers, to target key parameters that would help improve customer retention and future repurchase probability.

3.3 Review of Relevant Literature

Our work is related to research streams in online platforms, grocery retail, store switching, and choice modelling. Among existing studies, a majority of the earlier studies investigated customers' in-store shopping from a behavioral standpoint, using cross-sectional data. Leszczyc and Timmermans (1997) was the first study to have examined the temporal aspect of customers' store-shopping behavior (which they termed as *store-switching behavior*) using panel data. Following this, several studies have examined the drivers of store-switching in both online and offline services (Bell and Lattin, 1998; Leszczyc et al., 2000; Keaveney and Parthasarathy, 2001; Rhee and Bell, 2002; Bansal et al., 2005; Sloot and Verhoef, 2008; Wiebach and Hildebrandt, 2012; Van Lin and Gijsbrechts, 2014; Melis et al., 2015; Su et al., 2016; Hult et al., 2019; Singh and Rosengren, 2020; Wieland, 2021; Richards and Liaukonytė, 2022; Wang et al., 2023), and channel choice (Chintagunta et al., 2012). Among the prominent studies listed, Bell and Lattin (1998) investigated the temporal nature of consumers' shopping behavior based on the store-price format, such as EDLP (every-day-low-price) vs HILO (high-low) strategy. Bansal et al. (2005) provide a comprehensive study on the drivers of consumers' switching behavior by utilizing a push-pull and mooring (PPM) framework. The study suggests that switching behavior is not only driven by factors related to the retail store and previous transactions (such as service quality issues that *push* the customer away), but also include factors such as the presence of competitors trying to *pull* the customers to switch stores, along with mooring factors, which are high/low switching costs that make it harder/easier to switch. With the growing competition in the online grocery space, both the push and the pull factors are likely to play a significant role in determining customers' switching behavior. Our current work, focuses on 'push'-based factors such as service quality issues, as well as 'pull'-based factors such as the presence of multiple store alternatives.

Among more recent studies, Singh and Rosengren (2020) adapted the PPM framework by Bansal et al. (2005) to an online grocery retail context, to develop an understanding of the drivers of online grocery shoppers' switching behaviors, and provide new insights regarding customer loyalty in online purchases. On a different purview, Hu et al. (2019) demonstrate that a switching effect can be driven by information-based social learning, and mechanisms related to network size. Another study by Hult et al. (2019) examines the antecedents and consequences of customer satisfaction. The study examines whether there are any purchase-channel differences (online vs offline) in the antecedents of satisfaction and customer loyalty. Sloot and Verhoef (2008) examine the impact of brand delisting (removing a brand from the product assortment to put price pressure on the manufacturer) on customers' store switching and brand switching behavior.

Our work on customers' store-switching (and platform-exit) behavior is most closely related to Rhee and Bell (2002), and Musalem et al. (2010). To the best of my knowledge, Rhee and Bell (2002) is the only study the examines a switch in customers' primary store affiliation. The study reveals that customers are less-likely to change the store of their primary affiliation due to temporary price promotions. The authors observe that a majority of switches occur across competing store chains of the same price format, and suggest that 'format loyalty' is an important aspect of customer behavior. Our work extends upon this study by focusing on customers' online 'preferred' store affiliation, while investigating how service quality parameters and customer-specific behaviors could affect the 'primary' store affiliation, and the decision to exit the online service.

A majority of existing studies have focused on the decision to switch or not to switch. However, a growing stream of literature has examined store switching as a choice model. Our study also examines customers' store choice in online grocery retail, and is therefore closely related to this field of studies. For example, Melis et al. (2015) examines the differences in consumers' shopping frequency and experience levels while shopping from online channels (as opposed to offline channels). The authors first identify the drivers of online store choice, and explore the factors that could lead to a change in such drivers, especially when multichannel consumers gain an online shopping experience. Our work differs from this study as our objective is to examine how online shopping experience interacts with store- or platform-specific quality issues to modify customers' intrinsic utility, and whether that could lead to a switch in the customers' preferred store brand, or an exit from the online grocery platform. Other prominent studies using a store-choice model include Bell et al. (1998), which is one of the first studies to have modelled store switching as a (linear utility) *store choice model*, with the basic premise that each customer is more likely to visit the store with the lowest total shopping cost. Leszczyc et al. (2000) estimate a dynamic hazard model to understand the factors influencing customers' purchase timing, and store choice. They further investigate competition at the store-brand level, based on pricing strategies such as EDLP and HILO.

Finally, our work is also related to existing literature on the effects of stockouts on consumers' response towards the service (or store) (Ezhil Kumar et al., 2021; Hoang and Breugelmans, 2023). For example, (Campo et al., 2000) develop a conceptual framework to integrate the determinants of differing consumer reactions to stockouts. They further demonstrate the relevance of their framework using survey data on out-of-stock (OOS) behaviors. Other studies have examined the impact of product and brand characteristics on customers' response to stockouts. For example, Sloot et al. (2005) examine the impact of brand equity and the hedonic level of the product on consumers' stockout responses. Our work also draws upon Musalem et al. (2010), where the authors develop a structural demand model to capture the effect of stockouts on customers' choice to either defer their purchase, or select a substitute product, or to not purchase at all. The authors build their model based on utility maximization principles, and demonstrate the consequences of out-of-stocks. Our work explores the effect of out-of-stocks in the context of a store choice (as opposed to a product choice). We argue that when faced with online stockouts, a customer has the option to either switch their preferred store in the next period (similar to the substitution option), or not to purchase from the platform again, i.e., platform-exit (which is similar to the no-purchase option).

3.4 Model and Methodology

We build upon the idea that although customers (users) may patronize multiple stores while shopping groceries online, they usually have a 'primary' store they are affiliated to

(Rhee and Bell, 2002), where the 'primary' online store brand is defined as the store-brand which receives the largest allocation of a user's online grocery expenditure in a given time period. Each discrete time period provides the user with an option to continue with the same online store-brand (i.e., users display inertia), or switch to a different alternative. Our objective is to model this transition behavior of users of online grocery platforms.

Customers purchasing from offline (brick-and-mortar) grocery stores often choose to remain loyal to a particular store (or store-brand) (Hansen and Singh, 2008; Puligadda et al., 2012) owing to benefits resulting from proximity to the store location, cognitive benefits such as familiarity to the store layout and product locations, and economic benefits from loyalty programs. However, such factors have little to no influence on online store choice because factors such as proximity, knowledge of store layout and product assortment, and store-specific loyalty programs, are either not applicable, or do not provide any distinct advantage over the alternative store-brands available on the platform. Therefore, it is unclear whether customers do have an online store of primary affiliation. Melis et al. (2015) demonstrate that often customers prefer to purchase online groceries from the same store-brand they use to purchase offline. This could mean that customers have inherent store-brand preferences, or primary allegiances, while purchasing groceries online. If that is true, then factors such as online price promotions and discounts are less likely to affect the change of one's primary online store affiliation, although it might cause a trip- or order-specific switch, where one switches to a different store brand for a single order. This frequency of switching can also vary depending on the nature of the user. Therefore, from a managerial perspective, it is important to determine whether the tendencies to switch a primary (online) store-brand are related to the behavioral aspects of the shopper, or explicit actions taken by either the store or the platform itself. In the following sub-sections, we describe two different approaches through which we examine the store-switching, and store choice of customers.

3.4.1 Bayesian Analysis of Users' Store-switching and Platform-exit

In this section, we describe the model specification for customers' online store-switching behavior, with the caveat that we examine store-switching of a customer's primary store of affiliation. Let us consider a set of customers (users), u, where $u\varepsilon(1,2,...U)$. The online platform displays a list of all grocery store brands, s, where $s\varepsilon(1,2,...S)$, to the user in no particular order (such as alphabetical, or popularity-based). The customers therefore have access to all store brands available in the city, which are enrolled with the platform, and are open when the customer places an order. Let $s_{u,t}$ be the 'primary' online store-brand of user u at time period t. The switching index $y_{u,t}$ is defined as:

$$y_{u,t} = 1 \text{ if } s_{u,t} \neq s_{u,t+1} - (1)$$

where, $y_{u,t}$ takes the value of 1 if the primary store selected by the user in the upcoming (next) time period is different from the focal store, and 0 otherwise.

In other words, we examine whether store- and platform-specific factors such as the online price, and quality parameters in the focal period could cause a switch in the primary store in the upcoming period. Next, we further consider the scenario where a customer does not make a purchase at all in the upcoming period. The *no-purchase* scenario is termed as a *platform-exit*, which indicates that the customer has exited (or stopped using) the service. However, it is possible that the customer did not make any purchase in a specific period due to unknown factors such as travel to a different city. Therefore, to better capture the *no-purchase* option, we examine whether the customer made a purchase in any one of the following two time-periods.

At this point, we remind the readers that since we want to observe customers' long-term affiliation to a store brand, we need to designate an appropriate duration of time as a 'time period'. Unlike purchases from an offline store, where a 1-week duration seems an appropriate period to observe customers' primary store affiliation, online purchases from a grocery platform are less-frequent and might take place few times a month. Table 3.1

demonstrates that on an average, customers in our focal platform place around 3 orders per month. We therefore define an appropriate time-period for long-term online store affiliation as 1 month. However, as mentioned above, to examine customers' platform (service)-exit behavior, we determine a 2-month period to be more appropriate. Therefore, our focal variable $switch_{u,t}$ = 1 if in the upcoming period, a customer either switches to a different primary store, or does not make any purchase for the upcoming two consecutive time periods. Later, we also try two different operationalizations: First, we define $y_{u,t}$ = 1 if $s_{u,t} \neq s_{u,t-1}$, i.e., $y_{u,t}$ takes the value of 1 if the primary store selected by the user in the current period is different from that of the previous period; and second, we only observe the instances of store-switching and designate the user-periods with a platform-exit with a missing value.

Using equation (1), we can define the probability of switching from the primary online store brand as: $Switch_{u,t}$ = $Pr(y_{u,t} = 1)$, and then use the standard link function, Φ, to associate the probability of a store-switch (in the next period) to observable and unobservable characteristics of the online platform, store, and user at each switching incident.

$$Switch_{u,t} = \Phi(\alpha_u + \beta p'_{u,t} + \gamma q'_{s,t} + \delta(p'_{u,t} * q'_{s,t})) - (2)$$

where, α_u represents the unobserved innate switching tendency of a user u, $p'_{u,t}$ is a vector of shopping behavior variables (such as prior online grocery purchase experience frequency and value) with β as the response parameter. Next, $q'_{s,t}$ is a vector of platform-, store-, and time-specific variables (such as the online price and service quality), which could affect the long-term utility derived from an online store-brand, therefore leading to a switch in the primary store-brand, and γ is the response parameter. Finally, $p'_{u,t} * q'_{s,t}$ is an interaction that examines the moderating effect of time-varying user behavior on the time-varying store- and platform-specific parameters.

Let $Rel\_Util_{u,s,t}$ represent the relative utility that a user u derives from switching the primary online store s, in time period t. Then, we can exhibit a link between the discrete-time online store-choice and the long-term utility derived from changing the primary (online) store of affiliation as:

$$Rel\_Util_{u,s,t} = \alpha_u + \beta p'_{u,t} + \gamma q'_{s,t} + \delta(p'_{u,t} * q'_{s,t}) + \varepsilon_{u,s,t} - (3)$$

where, $\varepsilon_{u,t}$ is i.i.d and assumed to have a normal distribution with 0 mean and unit standard deviation.

Model Formulation and Estimation

We examine two models: Model (1), which is our base model, examines the main effect of time-varying customer behaviors, and time-varying platform, and store-specific parameters. Next, Model (2), which is the interaction model examines how time-varying customer behaviors moderate the effect of platform- and store-specific quality and price parameters on the decision to either continue or switch the 'primary' store brand. We further bifurcate this model to separately examine the effect on store-switching, and platform-exit behavior. Lastly, we investigate whether the prior incident of a primary store-brand-switch can exacerbate the tendency to exit the service (platform-exit). Following Rhee and Bell (2002); Musalem et al. (2010), and Van Lin and Gijsbrechts (2014), we use a Bayesian parameter estimation to analyze customers' online store-switching behavior. Unlike a maximum likelihood approach, whose objective is to determine a point estimate for the parameters of interest, the objective of a Bayesian approach is to estimate the posterior distribution of the parameters from the observed data (Musalem et al., 2010). Bayesian models combine"prior knowledge about model parameters with evidence from data" (Balov, 2016). This approach is particularly useful in our scenario, as the decision to make a choice regarding switching (or not) can be better examined using prior distributions of the parameter(s) of interest, to form an idea regarding the posterior probability of the occurrence of the event (switch).

To conduct a Bayesian analysis, we first require the parameter(s) of interest, θ. Next, we require the prior distribution of θ, $\pi(\theta)$, which summarizes what is previously known about θ (and might therefore be subjective to the researcher's interest). Following this, we require a likelihood function, $L(y|\theta)$. The likelihood function provides us with the distribution of our data, y, given the parameter value θ. For example, in our model, our outcome variable examines a binary (1/0) switching behavior, and we might therefore use a Probit model. Finally, we have the posterior distribution, $P(\theta|y)$, which summarizes the information within the data, y, along with the prior distribution information, $\pi(\theta)$. The posterior therefore summarizes what is previously known about the parameter of interest θ after the data has been collected. The Bayes' rule for posterior distribution is as follows:

$$P(\theta|y) \propto L(\theta|y)\pi(\theta) \text{ -(4)}$$

The posterior distribution $P(\theta|y)$ provides a complete description of θ and Markov Chain Monte Carlo (MCMC) methods are used to simulate $P(\theta|y)$.

To conduct our analyses, we utilize different sampling techniques, such as the Metropolis-Hastings algorithm, and Gibbs sampling. We also use different priors, ranging from slightly informative priors to hierarchical priors for our models. We further assess the convergence of our model using 2 or 3 parallel Gibbs chains with different initial values, which allows us to gather Gelman-Rubin statistics for model convergence. In the empirical analysis section, section 5.1, we describe more about the choice of sampling technique, choice of priors, and assessment of model convergence.

In this section, we broadly described our estimation methodology when our objective is to investigate customers' store-switching and platform-exit behavior. In the following sub-section, we describe our approach when we want to investigate the explicit choice of customers when presented with multiple online store-chain alternatives. In particular, we utilize the random utility maximization framework along with a multinomial logit (MNL) choice model to examine factors that lead to choice of a particular online store brand.

3.4.2 Multinomial Logit - Online Store Choice Model

In this section, our primary interest is to model customers' (users') choice of an online store brand (chain). We utilize the random utility maximization framework, wherein users select a store (in this case, a store ”brand”), that provides them the highest utility. Following Chintagunta et al. (1991); Bell et al. (1998); Tang et al. (2001); Briesch et al. (2013), and Melis et al. (2015) we use a random-coefficients multinomial logit (MNL) model to analyze store brand choice in an online grocery platform. We assume that a user u selects an online store brand s while shopping from an online grocery platform, such that the sum of the utility derived from placing orders over a fixed period of time, t, $Utility_{ust}$, is maximum. The distinguishing feature of our online store-brand choice model is that we evaluate users' choice of a ‘primary' online store brand, defined as the store-brand that receives the highest allocation of user expenditures in a fixed time period, while shopping for groceries using an online platform (aggregator). We may therefore specify the utility of selecting a store, s, where, $s\varepsilon(1,2,...S)$ in time period t, as:

$$Utility_{ust} = \alpha_u + \beta_{ut}^{'} x_{st} + \xi_s + \eta_t + \varepsilon_{ust} - (5)$$

where, α_u represents time-invariant user demographics, and time-varying behaviors, such as the user's online grocery shopping experience and past spending. Next, x_{st} is a vector of platform-, store- and period-specific covariates, such as the online price and service quality, which could affect the long-term utility derived from an online store-brand. The vector of random coefficients $\beta_{ut}^{'}$ varies across users and time periods, and describes individual decisions and preferences across time periods. The term ξ_s represents store-specific controls such as whether the store-brand is a ‘partner' store, and η_t represents time-specific controls that could affect both the consumption patterns and the utility of customers, such as whether the shopping was performed during the Covid19 period. Finally, ε_{ust} is an individual-specific error term, modeled as an i.i.d. random variable from a normal distribution with 0 mean, and unit standard deviation.

We now describe the probability that an online store-brand s, is chosen as the primary store-brand by the user u, during time period t $(Pr(y_{ust} = 1))$ as:

$$Pr(y_{ust} = 1) = \frac{exp(Utility_{ust})}{\Sigma_{s=1}^{S} exp(Utility_{ust})} - (6)$$

where, each time period t may consist of multiple online orders, $j\varepsilon(1,2,..,n)$, placed by the user, u from the available store brands, $s\varepsilon(1,2,...S)$.

Unlike the selection of a primary offline store (brand), where factors such as distance to the store, or familiarity to the assortment or product locations would matter more to the user, the choice of an online 'primary' store-brand is unlikely to be affected by the aforementioned factors. Furthermore, the selection of a 'primary' online store is less likely to be affected by factors such as temporary price promotions, and discounts, which are common in online retail. We therefore classify factors that could affect users' utility in a long-term online store-switch (i.e., switching the store brand of primary affiliation) as follows - platform-specific factors; heterogeneity across store brands and other store-specific factors; and heterogeneity among users.

Among store-brand-specific factors, while both online prices (across categories) and online assortment size could be essential in determining long-term affiliation with an online store-brand (Briesch et al., 2013), online assortments can vary across time periods, and even across orders within a time period. Although, we have information regarding the availability of brands across categories in a given store, we cannot be certain whether all the brands were displayed to a user in a given time period. Therefore, in our estimation, we consider category-wise store- and period-specific online prices, to account for *store-specific factors* that may affect users' utility. Among *platform-specific factors*, we consider service quality parameters such as missing (stocked-out) items, substitutions (for stockouts), and delays, as possible factors that might affect users' utility. Finally, to account for *heterogeneity among users*, we include users' online experience, and store-brand-specific online experience, along with other user demographics such as age, and gender, in our model. Therefore, equation (5), which describes the utility derived from a primary store-brand in a given time period may be expanded as follows:

$$Utility_{ust} = [\alpha_u + \alpha_1 Experience_{ut}] + [\beta_1 missing_{st} + \beta_2 delay_{st} + \beta_3 substitutions_{st}] + [\gamma_4 price_{st}] +$$
$$[\delta_1 Experience_{ut} * missing_{st} + \delta_2 Experience_{ut} * substitutions_{st} * + \delta_4 missing_{st} * substitutions_{st} +$$
$$\delta_5 Experience_{ut} * missing_{st} * substitutions_{st}] + \xi_s + \eta_t + \varepsilon_{ust} - (7)$$

where, ξZ_{st} is the set of additional control variables, and ε_{ust} is the error term.

The first square brackets indicates customer-specific factors that could impact utility derived within a time period from online shopping experiences. The second and third square brackets indicate platform-specific service quality parameters, and store-specific pricing factors that could affect users' utility. The last brackets indicate the shift in online utility as a result of growing online experience. Specifically, we capture the moderating effect of prior online shopping experience (and store-specific online experience) by including interaction terms between $Experience_{ut}$ and service quality parameters - $missing_{ust}$, and $substitutions_{ust}$ to examine whether prior online experience could mitigate the possible detrimental effects of stockouts. Lastly, we account for unobserved heterogeneity by making the main and interaction effects *random* over users.

We describe the results of the analyses in the empirical analysis section 5.2. In the following sub-section, we begin with describing the setting and data.

3.5 Setting & Data

We use data on user purchases from 98 store brands, consisting of 1380 stores, across Italy. No city contains more than 35 unique store brands. A discussion with the company revealed that the large number of store brands is because some store brands are actually non-grocery stores such as a bakery shop, or a stationery shop that also sells some specific grocery items etc. This provides further reason for us to focus on ‘primary’ stores as the main unit of analysis for examining store-switching, or store-choice behavior of the customers. On an average customers switch between 2 to 3 ‘primary’ store brands, but might switch brands more frequently, on a per-order basis.

For the analysis on store-switching, we organize our panel data to have a unique identifier called “user-period” because we want to focus on customers’ (users’) ‘primary’ online store-brand switch. Specifically, we have one primary store in each time period for each user ($main\_store$). Since customers might not purchase in a couple of periods, but come back later, we also include those periods in which a customer did not make a purchase and indicate the $main\_store$ with a missing value, therefore excluding them from our analyses. Our data consists online grocery purchases across 802,235 non-missing user-periods (by 118,929 unique users) from 1^{st} Jan 2019 to 1^{st} October 2021. As mentioned previously, we define an appropriate time-period to observe long-term (primary) store switching as 1 month. Therefore, our dataset consists of 33 time periods. For each unique user-period, we identify the $main\_store$ as the store-brand which receives the highest allocation of user expenditures in that time period. Table 3.1 provides descriptive statistics of the users’ online purchase (such as frequency, purchase value etc.), along with the statistics on selected model covariates.

Table 3.1: Descriptive Statistics

Variable Description	Mean	SD	Min	Max
Shopping frequency (#orders/period)	3.13	2.24	1.00	106.00
Spend per order (in ε)	77.72	41.94	1.39	1329.58
Overall spending (per user, in ε per period)	178.28	168.23	4.57	7329.66
Brands (Chains) frequented in a period	1.32	0.60	1.00	9.00
Brands (Chains) frequented overall	2.05	1.16	1.00	16.00
Experience	27.06	50.31	1.00	1300.00
Missing items	1.45	2.24	0.00	92.00
Substituted items	1.89	2.46	0.00	68.00
Delay	7.83	19.36	0.00	111.41

We define $switch$=1 in each non-missing time-period if the primary store in the focal time period is different from the primary store in the next time period, or the customer does not make an online purchase in the two consecutive upcoming time periods; and 0 otherwise. Since we do not observe any purchases beyond the 33^{rd} time-period, we indicate $switch$ as missing for the last observed period. To ensure robustness of our estimates, we an additional variable, $platform\_exit$, which equals 1 if the customer does not make an online purchase in the two consecutive upcoming time periods; and 0 otherwise. The objective is to examine whether our parameters of interest might have a different effect on platform-exit as opposed to store-switching.

Next, we describe our model covariates. We select our model covariates based on our research question, which is to investigate the extent to which online price and quality parameters can affect the switching (choice) of customers' 'primary' store-brand, and whether such switching can have an indirect secon-order effect on customers' decision to exit the online platform (service). We split our covariates onto three broad categories: *price-specific* factors; *user-specific* factors; and *service quality* factors. The first category includes a variable on the total value of purchase (see definition in Table 3.2) made by a user in the focal time period (*purchase_val*). User-dependent covariates include the online purchase experience of the user (*experience*), user demographics such as age and gender, and an identifier as to whether the customer moved to a different city during the estimation period ($diff\_city$). Lastly, the service quality covariates include the average number of stockouts (*missing_store*), substitutions (*substitutions_store*) and delays (*delay_store*) that the user experienced in the focal period, and an identifier to determine whether the service quality was poor (for e.g., higher stockouts) because the period fell under the first-wave of Covid19 (i.e., March 2020-June 2020; is_covid). Table 3.2 provides a description of the operationalization of the variables used in our analyses.

For the analysis on store-choice, to perform an MNL, we need to examine cities sepa-

Table 3.2: Variable Operationalization

Variable	Description	Notation
Switch or Exit	1 if the user's primary store affiliation changes in the next period, or the user exits the service in the next period	switch
Platform-exit	1 if the user exits the service (assuming a 2-period cut)	platform_exit2
Switch store	1 if the user's primary store affiliation changes in the next period, and missing if the user exits the service	switch_n
User category share	Category share of user *u* for each of the 22 categories of online grocery purchases from the platform during the estimation period	$s_{u,c}$
Delay	Average delay encountered when the order was purchased from the primary store (standardized)	delay
Value of purchase	Summation over 22 categories of the avg. unit price across all SKUs within a category for each 1-month time-period, weighted by the user's category share (standardized)	$purchase\_val_{u,p} = \sum_{c=1}^{22} s_{u,c} * price_{c,p}$
Experience	Frequency of purchases made from the online platform before the focal period (standardized)	experience
Missing (stockouts)	Sum over 22 categories of the avg. no. of stockouts across all SKUs within a category for each time-period, weighted using the user's category share (standardized)	$missing_{u,p} = \sum_{c=1}^{22} s_{u,c} * missing_{c,s,p}$
Substitution (of stockouts)	Sum over 22 categories of the avg. no.of stockout-based substitutions across all SKUs within a category for each time-period, weighted using the user's category share (standardized)	$substitutions_{u,p} = \sum_{c=1}^{22} s_{u,c} * substitution_{c,s,p}$

rately, as the most commonly used store brands might differ among cities. We focus our analyses on 3 cities in Italy (Milano, Roma, and Torino) that have the largest presence of our focal online grocery platform. The combined purchases from these three cities amount to ∼57% of the overall purchases during our estimation period. We observe that in each of these cities, customers have access to different store brands. We create three separate datasets, one for each city, where we observe customers' store choice in each period. The city of Milano consists of 161,120 user-periods (consisting of 26,199 users and 10 store brands), Roma consists of 160,930 user-periods (consisting of 23,470 users and 10 store brands), and Torino consists of 129,402 user-periods (consisting of 17,500 users and 7 store brands). Similar to our analysis on store-switching, we focus on the choice of a 'primary' (long-term) store in each time period.

Precisely, our objective is to investigate what factors could influence customers' selection of online store-brand "A" over store-brand "B" as their preferred primary store, in the focal time-period, when presented with multiple options that require limited to no switching costs. The variables of interest are similar to those of the store-switching model.

3.6 Empirical Analyses

3.6.1 Analyses on Store-Switching and Platform-Exit

Analyses on Store-Switching

Extending eqn. (3), we consider the following model :

$$Switch_{ust} =$$
$$\beta_0 + \beta_1 experience_{ut} + \beta_2 price_{st} + \beta_3 missing_{st} + \beta_4 substitutions_{st} + \beta_5 delay_{st} + u_i + \varepsilon_{ust}$$

where, u_i is the random effect for users.

For the purposes of comparison, we first use a random-intercept mixed-effects probit estimation to fit our model using maximum likelihood (ML) estimation, which pools information across users and provides realistic predictions. Results of the estimation are displayed in Table 3.3.

Table 3.3: ME Probit model for users' store-switching

	(1)	(2)	(3)
purchase_val	-0.097***	-0.097***	-0.097***
	(0.004)	(0.004)	(0.004)
missing	0.129***	0.127***	0.130***
	(0.003)	(0.003)	(0.003)
substitutions	0.135***	0.132***	0.133***
	(0.003)	(0.003)	(0.003)
experience	-0.141***	-0.150***	-0.151***
	(0.003)	(0.003)	(0.003)
delay	0.017***	0.017***	0.019***
	(0.002)	(0.002)	(0.002)
missing*substitutions		-0.004	-0.002
		(0.003)	(0.003)
missing*experience		-0.011***	-0.011***
		(0.003)	(0.003)
substitutions*experience		-0.023***	-0.022***
		(0.003)	(0.003)
missing*substitutions*experience		0.004	0.003
		(0.004)	(0.004)
missing*delay			-0.011***
			(0.002)
delay*experience			-0.005
			(0.004)
missing*delay*experience			-0.000
			(0.003)
delay*substitutions			-0.003
			(0.002)
missing*delay*substitutions			-0.002
			(0.002)
delay*experience*substitutions			0.007*
			(0.004)
missing*delay*experience*substitutions			-0.004
			(0.003)
constant	0.079***	0.075***	0.076***
	(0.004)	(0.004)	(0.004)
var(_cons[user_id])	0.420***	0.419***	0.419***
	(0.005)	(0.005)	(0.005)
Observations	435,737	435,737	435,737

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Now, we consider the following Bayesian model (of the above equation) for these data:

$\varepsilon_{ust} \sim i.i.d.\ N(0;\sigma_0^2)$

$u_i \sim i.i.d.\ N(0;\sigma_u^2)$

$\beta_0 \sim cauchy(0,2.5)$

$\beta_{1to5} \sim cauchy(0,2.5)$

$\sigma_0^2 \sim InvGamma(0.01;0.01)$

$\sigma_u^2 \sim InvGamma(0.001;0.001)$

The main parameters of interest are the regression coefficients β_{0-5} and the variance components σ_0^2 and σ_u^2. The user's random effects, represented using u_i are called nuisance parameters. We use cauchy priors for the standardized regression coefficients following the approach suggested by Gelman et al. (2014), and use random effects and inverse-gamma priors for the variance parameters. Table 3.4 displays the results.

Table 3.4: MH Sampling Parameter Estimates: Store-switching

	(1)	(2)
purchase_val	-0.082**	-0.082**
	(0.003)	(0.003)
delay	0.019**	0.019**
	(0.002)	(0.002)
experience	-0.212**	-0.221**
	(0.002)	(0.002)
missing	0.139**	0.137**
	(0.002)	(0.002)
substitutions	0.140**	0.135**
	(0.002)	(0.002)
missing*experience		-0.012**
		(0.003)
substitutions*experience		-0.031**
		(0.003)
missing*substitutions		-0.009**
		(0.002)
missing*substitutions*experience		0.002
		(0.003)
constant	-0.079**	-0.083**
	(0.002)	(0.002)
Observations	435,737	435,737

** $p < 0.05$

We estimate the results based on the last 10000 posterior draws (MCMC simulation) from the MH-algorithm whose convergence is verified by the initial 2,500 ‘burn ins’ using the Gelman-Rubin statistic. We set a random-number seed to ensure replication of our results. Upon running the model, the analysis failed after the first 200 ”burn-ins” because mata failed to provide enough memory to estimate a matrix of the size 802,235 (which is the number of observations) × 118,934 (which is the number of parameters, including the random effects for users). We therefore conduct our analysis excluding the random effects for users. Table 3.4 displays the results of the estimation. Later, in the robustness checks we perform the same analysis, but conduct it only for the last 6 periods of our estimation sample, while reducing both the ”burn-ins” and the simulations. We have model convergence in this scenario and we present the results of this estimation in the Appendix Table C.15. The results of the estimation are similar in direction to those obtained excluding the random effects for user, but slightly different in magnitude of the coefficients.

The findings from the Bayes estimation, when compared to the frequentist (mixed effect ML estimation), are quite similar, despite the non-inclusion of the random coefficient. Both the magnitude and the direction of the coefficients of models 1 and 2 are similar across both tables. The only notable differences are the direction of the constant terms, which is quite likely due to the non-inclusion of the random coefficients. The significant parameter estimates based on the 95% posterior-band have been indicated. The model’s specification follows that the probability of switching increases with the effect of the covariate. We therefore read the estimates presented in the table as follows: On average, the probability of switching the primary online store is lower (higher) if the posterior mean of the associated parameter(s) is negative (positive). Specifically, we note that if the customers have higher-value purchases, and their prior experience of purchasing groceries online is higher, they tend to be less-likely to switch their primary (online) store-brand. On the other hand, the probability of switching is higher if there are service quality issues such as delays, and stockouts. While, we observe a positive main effect of substitutions on switching, it could be possible that the effect if partially due to stockouts, and not substitutions per se. We therefore conduct an analysis (presented in model 2 of Table 3.4) including interactions between missing items, substitutions, and prior experience. We now observe that when missing items are substituted, the switching probability is lower. Interestingly, we also observe that experience plays a significant role in mitigating the negative consequences of service quality on customers' tendency to switch their primary store.

We check whether there are any model convergence issues for both the models, and do not observe any noticeable issues. For e.g., the average efficiency of model 1 is $\sim$ 17%, which seems fine as efficiencies $\geq$ 10% are considered good for MH-algorithm, while those $\leq$ 1% are considered problematic, and might suggest non-convergence of the parameters. We further conduct a postestimation to check the effective sample sizes (ESS) and the efficiencies of the individual parameters and observe no issues. The acceptance rate is close to 0.44, which is also fine, since the MCMC simulation is assumed to be most effective when the acceptance range is between 0.2 (for high-dimensional problems) to 0.5 (for low-dimensional problems). Lastly, the Gelman–Rubin convergence diagnostic displays coefficients $<$ 1.1, which indicates convergence for the parameters.

Next, we plot the trace plot, the histogram, the autocorrelation plot, and the kernel density plot to determine possible issues regarding model convergence, autocorrelation. Figures 3.1 to 3.5 display the plots for the variables of interest: purchase_val, experience, delay, missing, and substitution, respectively. The trace plot for each of the five variables indicates that convergence was achieved. Next, the plots for autocorrelation demonstrate that autocorrelation is negligible after 30 periods for each of the five plots, but autocorrelation dies off quickly for Chain 1, where we set no initial value. Lastly, we check the histogram and the density plot and we find that the plots are in agreement with a normal distribution. In particular, the kernel density plot depicts that if the chains have mixed well (i.e., convergence was achieved), the 1^{st} half, the 2^{nd} half, and the overall plots must be close to each other, which is what we observe here. So, we can conclude that our models do not have any identifiable issues.

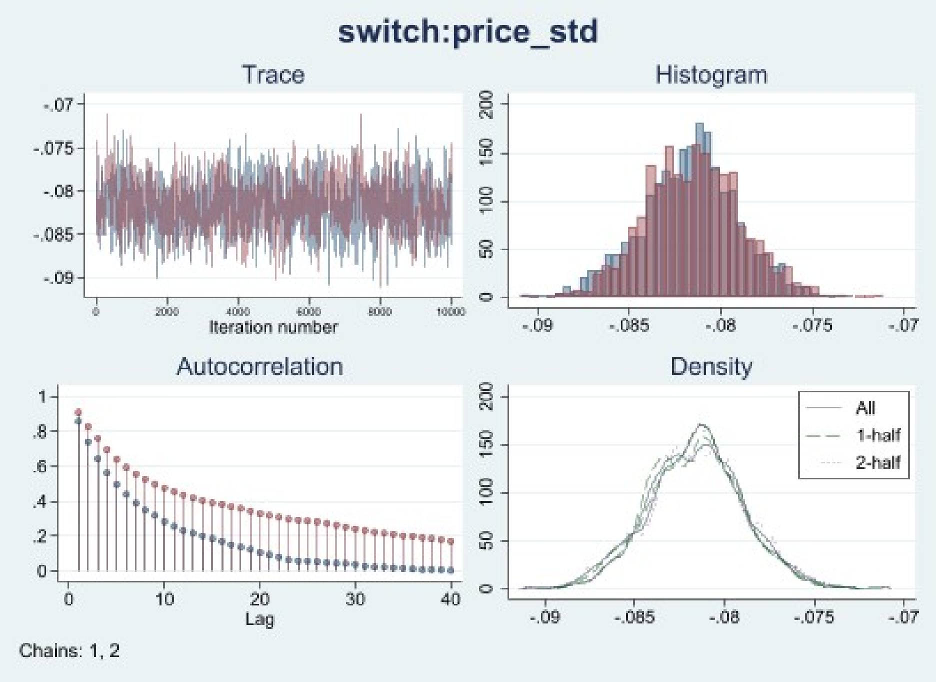


Figure 3.1: Plot for Switch:Purchase Value

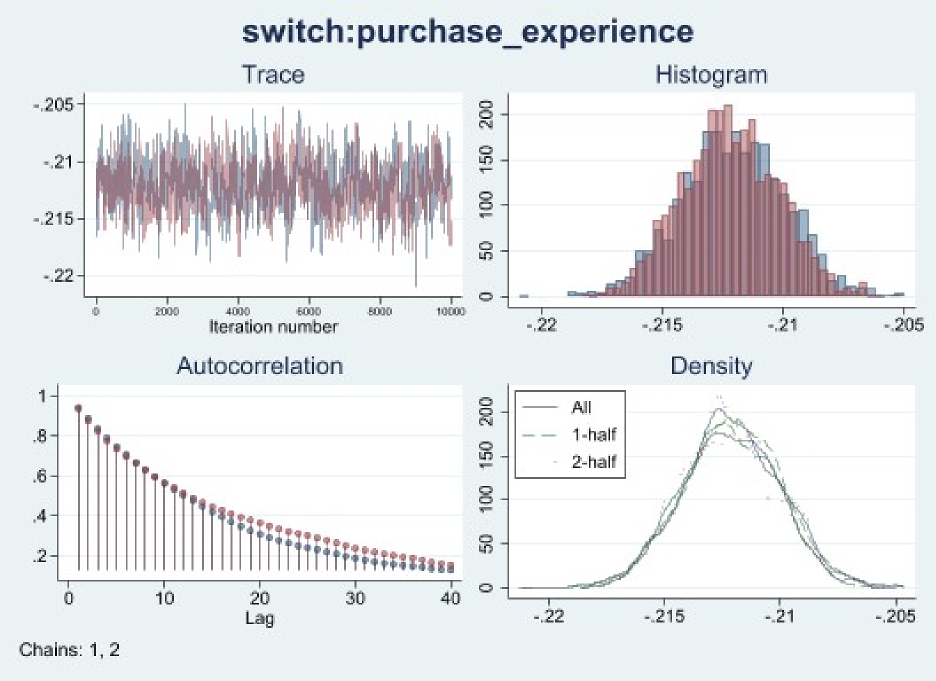


Figure 3.2: Plot for Switch:Experience

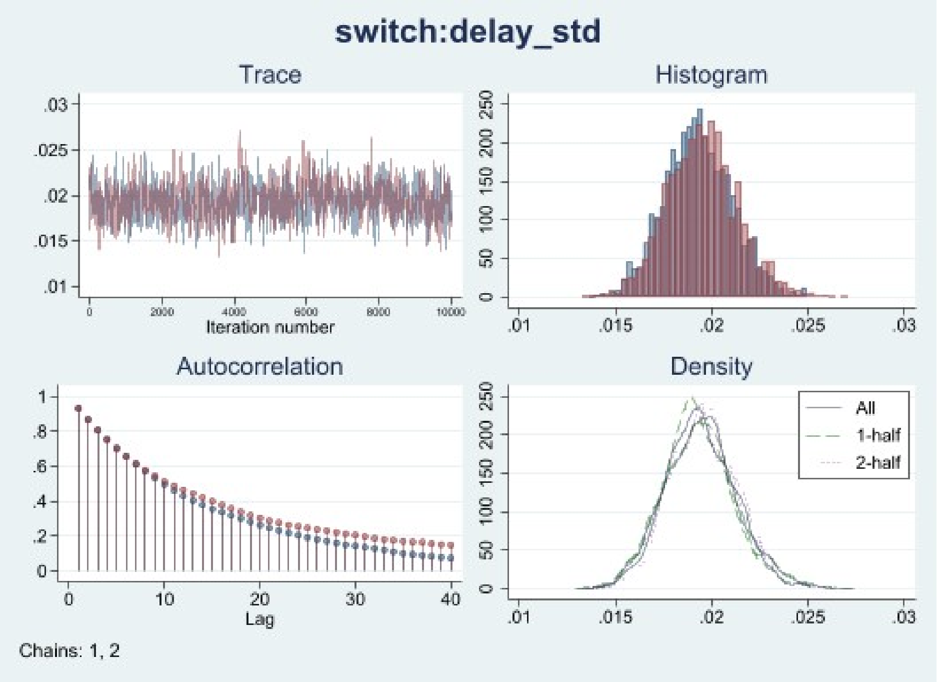


Figure 3.3: Plot for Switch:Delay

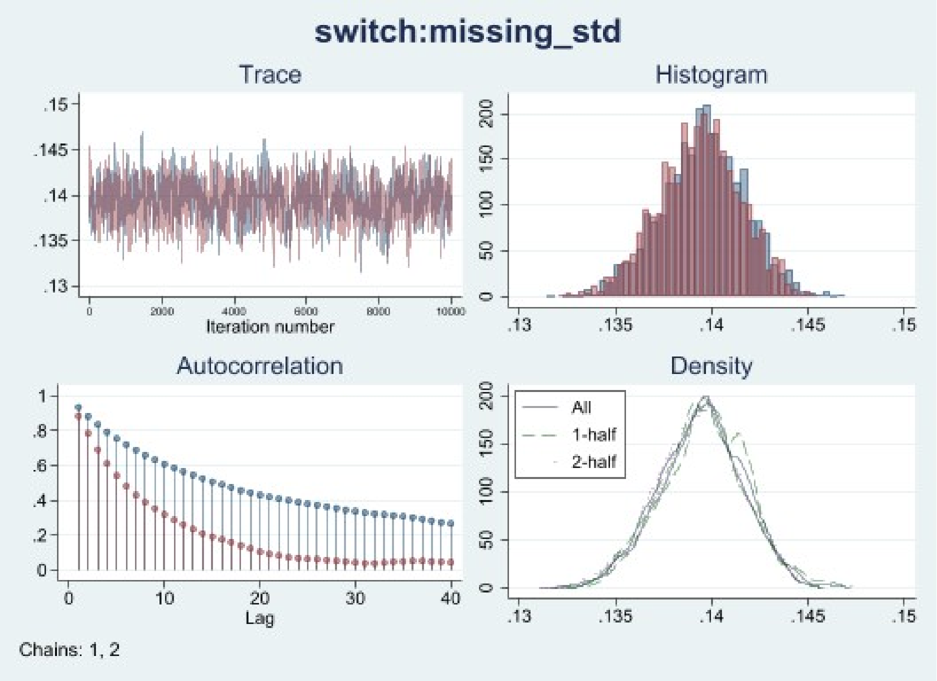


Figure 3.4: Plot for Switch:Missing

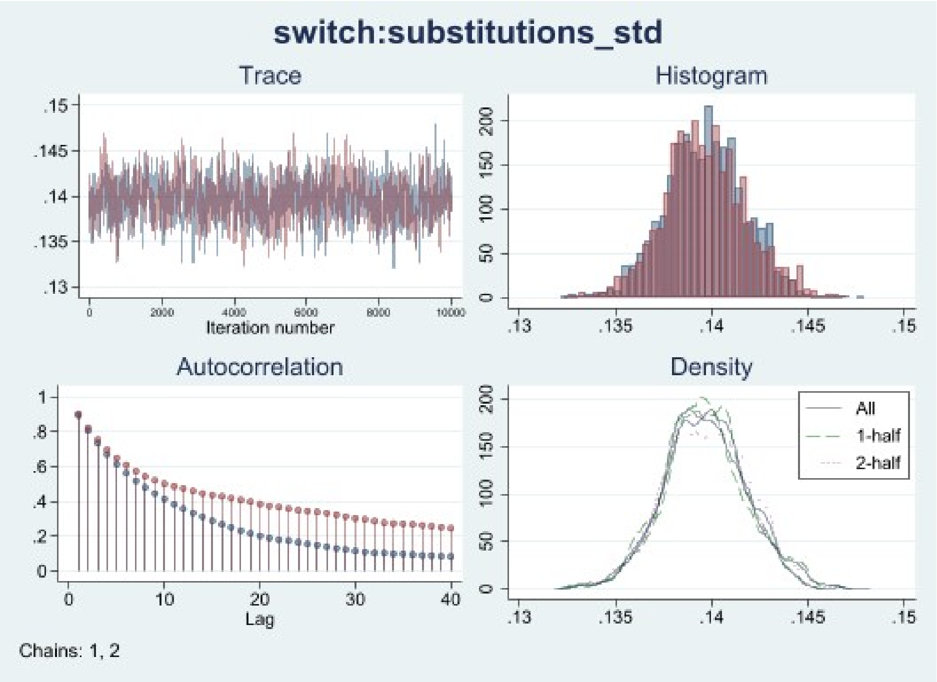


Figure 3.5: Plot for Switch:Substitutions

As an additional step, we perform model calibration and validation to determine the log predictive score of our model. We split our data after setting a random seed, into a calibration sample (80%) and a validation sample (20%) while clustering for user-period, and balancing the outcome variable, *switch*. Next, we compute the log predictive score of the main model, which is model (1) of Tables 3.3 and 3.4, and then the interaction model, model (2). Tables 3.5 displays the models' log predictive score comparison.

Table 3.5: Bayesian model log predictice score (LPS) comparison (Store-switching)

LPS	Mean	Minimum	Maximum
Model 1	0.874	0.196	49.012
Model 2	0.858	0.196	44.832

Notes: Using analytical posterior model probabilities (PMPs).

The results from Table 3.5 demonstrate that model 2 has the smallest mean log predictive-score (LPS), and since model 2's LPS of 0.858 is closer to the entropy value of 0.696, it is determined as the better model. The results are based on the 20% holdout sample. A full model with all interaction terms included (equivalent to model 3 of Table 3.3), yields an LPS of 0.695, which although closer to the entropy, has several issues such as multiple indicators with PIPs¡10%, and a low acceptance rate. Hence, we only consider models 1 and 2 as part of our estimation.

We further conduct an estimation using the MH-Gibbs sampling technique, the results of which are similar to what we observe here, and are displayed in Appendix Table C.12, Column 1.

We conduct the following additional robustness checks: (i) Including a control to identify whether the period coincides with the first wave of Covid19, which saw a sudden surge in online grocery ordering. The results of both the main and interaction models using the MH-estimator are presented in Appendix Table C.13 (ii) Including a control to identify if the customer changed her city, and a control to identify if the primary store was a partner store or not (which would imply that the platform had better visibility of the inventory availability of the items, and could therefore better address stockouts). Results are presented in Appendix Table C.14, Column 1 and Column 2, respectively (iii) Conducting analysis excluding the cases of platform-exit - i.e., platform exit is considered as missing. The results of the ME Probit estimation is presented in Table 3.16 of the Appendix, while the MH-sampling estimates are presented in Table 3.17. The plots of convergence and autocorrelation for all parameters are presented in Appendix Figures C.1 to C.5

Analyses on Platform-Exit

Next, we look at users' platform-exit behavior. In the previous sub-section, we examined the effect of our parameters of interest on users' probability to switch their primary stores, or exit the platform in the upcoming time periods. We further tested the robustness of our analysis by indicating the incidents of platform-exit with missing values (i.e, switch is neither equal to 0, nor 1, if there has been a platform exit).

In this section, we tease-out the effect of our parameters of interest on users' platform-exit behavior. Specifically, we denote all incidents where the customer exited the platform for two or more periods as 1, and all remaining incidents, where the customer did not exit but platform, but may or may not have switched their primary store as 0. Our objective is to determine whether user-specific behaviors, or platform- and store-specific factors could actually lead to a service exit.

While switching of the primary online retail store could have important implications for the grocery retailers enrolled with the platform, it may or may not have an effect on platform revenues. However, if the factors that lead to a store-switch could also lead to a platform-exit, then examining this effect could have important managerial implications for online grocery retailers. Following a similar approach (as in the section 5.1.1), we first use a random-intercept mixed-effects probit estimation to fit our model using ML estimation.

Table 3.6 below depicts the results of the estimates (for platform-exit) from the mixed-effects (ME) probit model.

Table 3.6: ME Probit model for Platform-exit (2-month)

	(1)	(2)	(3)
purchase_val	-0.209***	-0.206***	-0.206***
	(0.007)	(0.007)	(0.007)
missing	0.138***	0.181***	0.180***
	(0.004)	(0.006)	(0.006)
substitutions	0.132***	0.164***	0.160***
	(0.004)	(0.006)	(0.006)
experience	-1.035***	-1.015***	-1.023***
	(0.012)	(0.012)	(0.013)
delay	0.014***	0.015***	0.053***
	(0.004)	(0.004)	(0.006)
missing*substitutions		0.002	0.003
		(0.006)	(0.006)
missing*experience		0.111***	0.097***
		(0.012)	(0.012)
substitutions*experience		0.083***	0.067***
		(0.011)	(0.012)
missing*substitutions*experience		0.024**	0.024**
		(0.011)	(0.012)
missing*delay			-0.012***
			(0.004)
delay*experience			0.082***
			(0.011)
missing*delay*experience			0.004
			(0.008)
delay*substitutions			0.016***
			(0.005)
missing*delay*substitutions			-0.007*
			(0.004)
delay*experience*substitutions			0.051***
			(0.010)
missing*delay*experience*substitutions			-0.016**
			(0.007)
constant	-1.642***	-1.627***	-1.629***
	(0.006)	(0.006)	(0.006)
var(cons[user id])	-0.635***	-0.627***	-0.624***
	(0.028)	(0.028)	(0.028)
Observations	448,537	448,537	448,537

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Next, we consider the Bayesian model and generate parameter estimates for platform-exit. Table 3.7 displays the Metropolis-Hastings Parameter Estimates for Platform-Exit (using a 2-month cut).

Table 3.7: MH Sampling Parameter Estimates: Platform-Exit (2-month cut)

	(1)	(2)
purchase_val	-0.098**	-0.097**
	(0.003)	(0.003)
delay	0.012**	0.012**
	(0.002)	(0.002)
experience	-0.622**	-0.618**
	(0.005)	(0.005)
missing	0.085**	0.093**
	(0.002)	(0.003)
substitutions	0.079**	0.088**
	(0.003)	(0.002)
missing*experience		0.018**
		(0.005)
substitutions*experience		0.022**
		(0.005)
missing*substitutions		-0.000
		(0.003)
missing*substitutions*experience		0.014**
		(0.006)
constant	-0.980**	-0.977**
	(0.003)	(0.003)
Observations	448,537	448,537

** $p < 0.05$

Similar to the results of our analysis on customers' store-switching, we note that if customers have higher-value purchases, and their prior experience of purchasing groceries online is higher, they tend to be less-likely to exit the platform (service). On similar line, we also observe that the probability of exit is higher if there are service quality issues such as delays, and stockouts. We further conduct an analysis including interactions between missing items, substitutions, and prior experience (presented in model 2 of Table 3.4). Interestingly, we observe that prior experience of purchasing from the online grocery platform, plays a role in customers' learning. Specifically, we observe that while experience mitigates the detrimental consequences of service quality on customers' tendency to switch their primary store, it might exacerbate customers' tendency to exit the platform (service). This effect could be because experienced customers realize that the service quality issues are possibly not a factor of the store, and rather that of the platform. Moreover, experienced customers might be more sticky in their online store-brand preferences, and therefore prefer to exit the service in case of repeated service quality issues.

Table 3.8 displays the Log predictive score (LPS) comparison for the main and interaction models

Table 3.8: Bayesian model LPS comparison: Platform-exit (2-month cut)

LPS	Mean	Minimum	Maximum
Model 1	0.592	-0.036	37.263
Model 2	0.574	-0.037	32.688

Notes: Using analytical posterior model probabilities (PMPs).

We further conduct an estimation using the MH-Gibbs sampling technique, the results of which are similar to what we observe here, and are displayed in Appendix Table C.12, Column 2.

We conduct the following additional robustness checks: (i) Conducting analysis using a 1-month cut for Platform-exit. The results of the MH-sampling estimates are presented in Table 3.18 (ii) Including a control to identify whether a recent store-switch took place or not, to identify whether a recent store switch could further the cause of a platform-exit. The results of both the main and interaction models using the MH-estimator (2-month cut) are presented in Appendix Table C.19.

The convergence and autocorrelation plots for platform-exit are depicted in Figures 3.6-3.10.

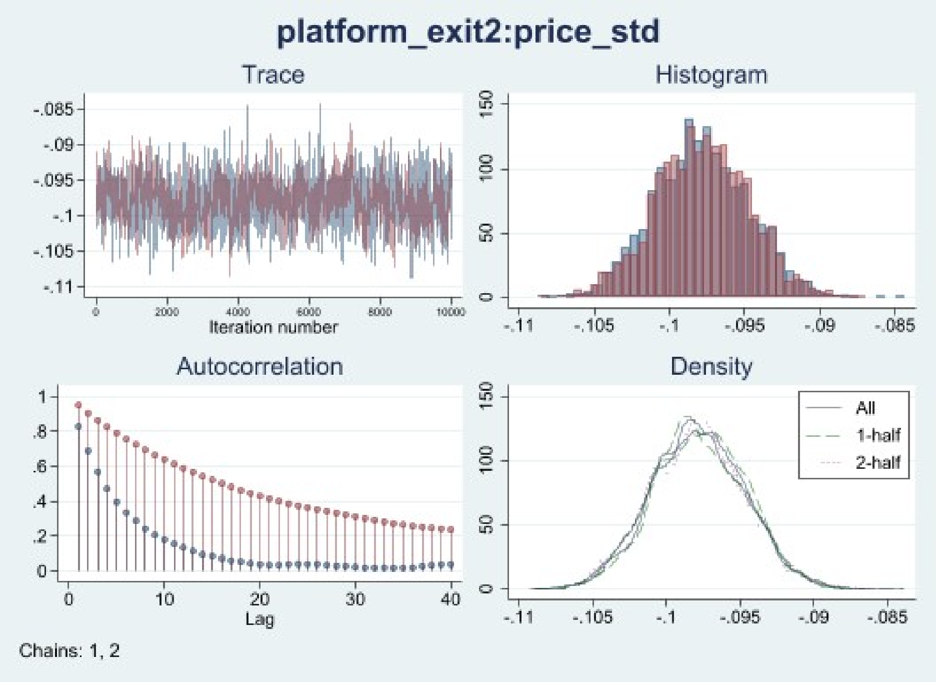


Figure 3.6: Plot for Platform-exit (2-month cut):Purchase Value

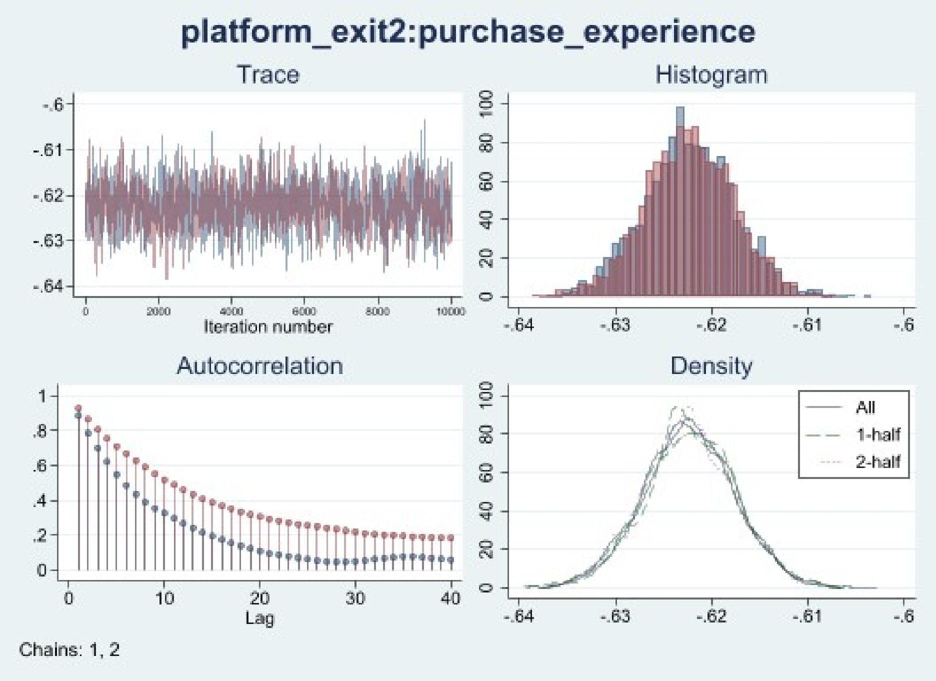


Figure 3.7: Plot for Platform-exit (2-month cut):Experience

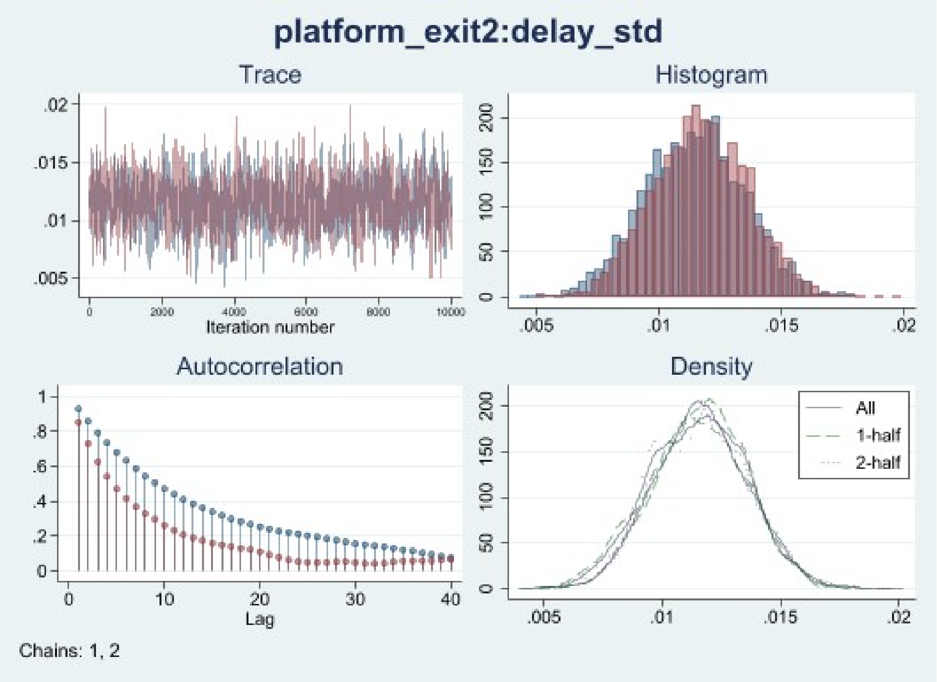


Figure 3.8: Plot for Platform-exit (2-month cut):Delay

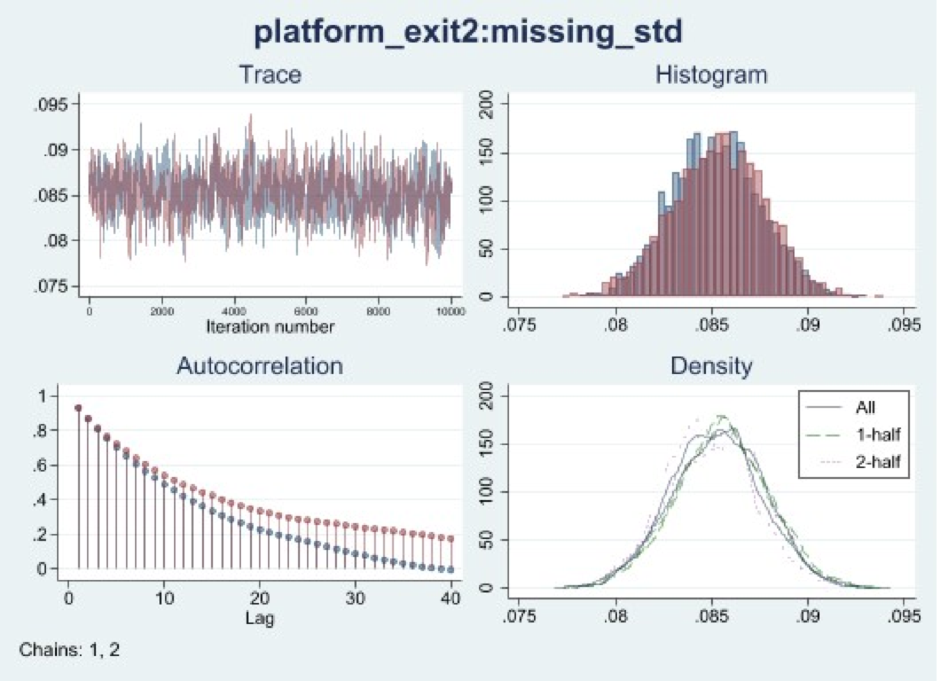


Figure 3.9: Plot for Platform-exit (2-month cut):Missing

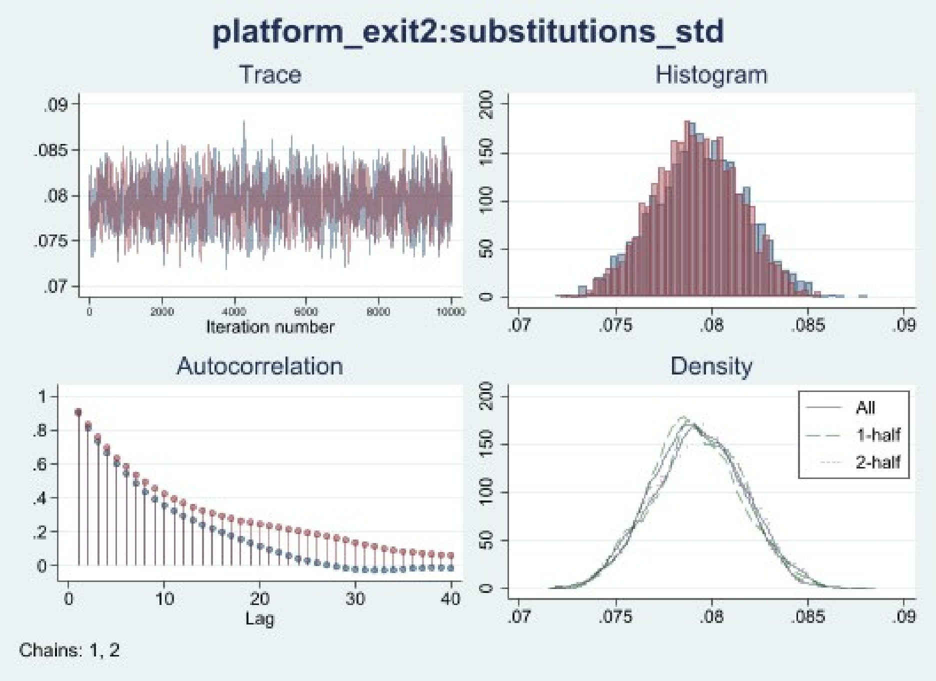


Figure 3.10: Plot for Platform-exit (2-month cut):Substitutions

The trace plot for each of the five variables indicates that convergence was achieved. Next, the plots for autocorrelation demonstrate that autocorrelation is negligible after 30 periods for each of the five plots, but autocorrelation dies off quickly for Chain 1, where we set no initial value. Lastly, we check the histogram and the density plot and we find that the plots are in agreement with a normal distribution. The kernel density plot depicts that if the chains have mixed well (i.e., convergence was achieved), the 1^{st} half, the 2^{nd} half, and the overall plots must be close to each other, which is what we observe here. We conclude that our models do not have any major issues.

3.6.2 Analyses on Store-Switching

In this section, we examine users' online store choice behavior in three cities of Italy where the presence of our online grocery platform is the largest ($\sim$ 57% of the total orders). As explained previously, since the customers in each city have access to a different set of online store brands, conducting the analysis at a city-level makes sense.

Our analyses focuses on the choice of users' primary store brand in a given time-period. Table 3.9 displays the relative risk ratios (RRR) for online store choice in the city of Roma (Rome). There are 10 primary store-brands in Rome, and we consider store-brand 57 as the reference (base) store. We display the coefficients for the three largest store brands in the table. However, the coefficients of the remaining 6 stores are also available. The choice of the 3 stores displayed, along with the reference store is based on the frequency of purchase. Since the coefficients displayed are the RRR, a value > 1 is considered to be better than the reference store, while a value < 1 is inferior.

Table 3.9: Multinomial Logit model: Roma

	(1)	(2)	(3)	(4)	(5)	(6)
9						
purchase_val	0.359***			0.351***	0.379***	0.347***
	(0.005)			(0.008)	(0.007)	(0.009)
experience	1.029***	0.811***	0.817***	1.068***	1.083***	0.748***
	(0.009)	(0.010)	(0.009)	(0.012)	(0.011)	(0.016)
missing		0.659***		0.699***		0.801***
		(0.011)		(0.011)		(0.015)
substitutions		0.276***	0.446***	0.298***	0.450***	0.319***
		(0.005)	(0.010)	(0.006)	(0.010)	(0.006)
delay		0.999	1.003	0.997	1.004	0.997
		(0.011)	(0.010)	(0.011)	(0.010)	(0.011)
stockout			0.596***		0.644***	
			(0.012)		(0.013)	
missing*substitutions						1.435***
						(0.035)
missing*experience						1.008
						(0.028)
substitutions*experience						0.513***
						(0.015)
missing*substitutions*experience						1.308***
						(0.049)
50						
purchase_val	0.406***			0.329***	0.357***	0.325***
	(0.007)			(0.008)	(0.007)	(0.008)
experience	0.915***	0.637***	0.644***	0.933***	0.942***	0.887***

(Cont'd on following page)

Table 3.9, continued

	(1)	(2)	(3)	(4)	(5)	(6)
	(0.011)	(0.009)	(0.008)	(0.013)	(0.012)	(0.015)
missing		1.007		1.076***		1.100***
		(0.013)		(0.014)		(0.016)
substitutions		0.528***	0.504***	0.568***	0.505***	0.563***
		(0.008)	(0.009)	(0.009)	(0.009)	(0.009)
delay		0.943***	0.952***	0.940***	0.951***	0.938***
		(0.010)	(0.009)	(0.010)	(0.009)	(0.010)
stockout			1.128***		1.234***	
			(0.018)		(0.020)	
missing*substitutions						1.109***
						(0.020)
missing*experience						1.130***
						(0.021)
substitutions*experience						0.738***
						(0.017)
missing*substitutions*experience						1.138***
						(0.028)
51						
purchase_val	1.629***			1.989***	1.886***	1.990***
	(0.016)			(0.030)	(0.024)	(0.030)
experience	0.831***	1.042***	1.020***	0.862***	0.835***	0.854***
	(0.007)	(0.008)	(0.007)	(0.008)	(0.007)	(0.009)
missing		0.983		0.931***		0.944***
		(0.011)		(0.011)		(0.011)
substitutions		0.893***	1.015	0.846***	1.000	0.864***
		(0.010)	(0.015)	(0.010)	(0.015)	(0.010)
delay		0.972***	0.970***	0.970***	0.970***	0.969***
		(0.009)	(0.008)	(0.009)	(0.008)	(0.009)
stockout			0.834***		0.792***	
			(0.012)		(0.012)	
missing*substitutions						1.119***
						(0.014)
missing*experience						0.949***
						(0.012)
substitutions*experience						0.970**
						(0.012)
missing*substitutions*experience						1.130***
						(0.018)
Observations	154239	88312	119115	88312	119115	88312
Pseudo R2	0.044	0.059	0.056	0.095	0.090	0.102

We first check the constant terms of the last model (model 6) for each of the stores (not displayed in Table), as our objective is to determine which store has the largest intrinsic utility holding all other parameters constant. Based on the significant values of the mean and parameter estimates, Store 51 experiences the highest intrinsic store utility (α Store51 = 1.016), relative to the base (reference) store 57, p $\leq$ 0.1), but has a decrease in intrinsic utility as consumers gather more online grocery shopping experience (α Store51 = 0.854, p $\leq$ .01). We further observe that as missing items are substituted, the intrinsic utility increases (> 1) for all 3 stores. However, in case of stockouts (missing items), experienced users show a significant positive utility only for store 50, but not for the remaining stores. Lastly, the triple interaction term of missing items (stockouts), substitutions, and customer experience displays a positive utility across all 3 stores. Overall, we can say that experienced users are more likely to loose intrinsic utility derived from their primary online store-brand, especially in case of stockouts (missing items). However, their utility increases if the missing items are substituted.

Table 3.10: Multinomial Logit model: Milano

	(1)	(2)	(3)	(4)	(5)	(6)
49						
purchase_val	0.988			0.937^{***}	0.974	0.942^{**}
	(0.013)			(0.023)	(0.018)	(0.023)
experience	0.865^{***}	0.897^{***}	0.891^{***}	0.896^{***}	0.882^{***}	0.889^{***}
	(0.010)	(0.011)	(0.010)	(0.013)	(0.011)	(0.014)
missing		1.068^{***}		1.073^{***}		1.068^{***}
		(0.020)		(0.020)		(0.021)
substitutions		1.032^{*}	0.969	1.038^{*}	0.970	1.037^{*}
		(0.020)	(0.022)	(0.020)	(0.022)	(0.020)
delay		0.901^{***}	0.886^{***}	0.901^{***}	0.885^{***}	0.901^{***}
		(0.012)	(0.010)	(0.012)	(0.010)	(0.012)
stockout			1.083^{***}		1.085^{***}	
			(0.025)		(0.024)	
missing*substitutions						1.056^{***}
						(0.022)
missing*experience						1.023

(Cont'd on following page)

Table 3.10, continued

	(1)	(2)	(3)	(4)	(5)	(6)
						(0.018)
substitutions*experience						0.962**
						(0.018)
missing*substitutions*experience						1.035*
						(0.022)
50						
purchase_val	0.168***			0.097***	0.119***	0.098***
	(0.003)			(0.002)	(0.002)	(0.003)
experience	1.295***	0.817***	0.840***	1.266***	1.304***	1.255***
	(0.011)	(0.007)	(0.006)	(0.013)	(0.012)	(0.014)
missing		1.163***		1.282***		1.254***
		(0.015)		(0.018)		(0.018)
substitutions		0.832***	0.638***	0.953***	0.666***	0.924***
		(0.012)	(0.011)	(0.014)	(0.012)	(0.014)
delay		0.793***	0.793***	0.797***	0.797***	0.798***
		(0.008)	(0.007)	(0.008)	(0.007)	(0.008)
stockout			1.455***		1.632***	
			(0.023)		(0.027)	
missing*substitutions						0.972*
						(0.016)
missing*experience						1.060***
						(0.015)
substitutions*experience						0.886***
						(0.013)
missing*substitutions*experience						0.909***
						(0.017)
51						
purchase_val	1.105***			1.133***	1.154***	1.145***
	(0.013)			(0.027)	(0.020)	(0.027)
experience	0.638***	0.736***	0.713***	0.681***	0.651***	0.683***
	(0.008)	(0.011)	(0.009)	(0.011)	(0.010)	(0.012)
missing		1.284***		1.271***		1.234***
		(0.022)		(0.022)		(0.021)
substitutions		1.395***	1.077***	1.384***	1.072***	1.346***
		(0.024)	(0.022)	(0.023)	(0.022)	(0.023)
delay		0.814***	0.800***	0.814***	0.802***	0.812***
		(0.011)	(0.010)	(0.011)	(0.010)	(0.011)
stockout			1.381***		1.365***	
			(0.029)		(0.028)	
missing*substitutions						1.023

(Cont'd on following page)

Table 3.10, continued

	(1)	(2)	(3)	(4)	(5)	(6)
						(0.018)
missing*experience						1.007
						(0.020)
substitutions*experience						1.089^{***}
						(0.021)
missing*substitutions*experience						1.032
						(0.020)
Observations	158142	86820	118790	86820	118790	86820
Pseudo R2	0.051	0.038	0.035	0.094	0.085	0.098

Next, we look at estimates of users' store choice in the city of Milan. Based on the significant values of the mean and parameter estimates, Store 50 experiences the highest intrinsic store utility (α Store50 = 2.041), relative to the base store 48, p $\leq$ 0.01), but also has the largest decrease in intrinsic utility as the purchase value of the order basket increases (α Store50 = 0.098, p $\leq$ .01). Store 51 also experiences the strongest decrease in intrinsic utility when missing items are substituted (α Store50 = 0.972, p $\leq$ .01). Lastly, the triple interaction term indicates that the intrinsic utility is also the lowest for store 50 in case of stockout-based substitutions when customer experience is higher. This effect can be explained by the tendency of 'inexperienced' customers to select the largest and most popular store brand as their default primary store. However, as their online experience increases, their tolerance for poor service quality decreases. Moreover, their intrinsic utility decreases even further for high-value purchases.

Lastly, we look at estimates of users' store choice in the city of Turin.

Table 3.11: Multinomial Logit model: Torino

	(1)	(2)	(3)	(4)	(5)	(6)
2						
purchase_val	4.377^{***}			6.686^{***}	6.073^{***}	6.732^{***}
	(0.074)			(0.173)	(0.132)	(0.175)

(Cont'd on following page)

Table 3.11, continued

	(1)	(2)	(3)	(4)	(5)	(6)
experience	0.840***	1.301***	1.270***	0.813***	0.791***	0.828***
	(0.009)	(0.015)	(0.013)	(0.011)	(0.010)	(0.012)
missing		1.111***		1.007		0.998
		(0.015)		(0.014)		(0.014)
substitutions		1.625***	1.573***	1.505***	1.558***	1.527***
		(0.021)	(0.025)	(0.021)	(0.025)	(0.021)
delay		1.081***	1.084***	1.093***	1.093***	1.090***
		(0.013)	(0.011)	(0.013)	(0.012)	(0.013)
stockout			0.992		0.897***	
			(0.015)		(0.014)	
missing*substitutions						0.960***
						(0.013)
missing*experience						0.936***
						(0.015)
substitutions*experience						1.267***
						(0.020)
missing*substitutions*experience						0.983
						(0.018)
5						
purchase_val	2.690***			3.230***	3.236***	3.240***
	(0.044)			(0.079)	(0.066)	(0.079)
experience	0.926***	1.236***	1.226***	0.944***	0.916***	0.924***
	(0.008)	(0.012)	(0.011)	(0.010)	(0.009)	(0.012)
missing		0.717***		0.684***		0.680***
		(0.009)		(0.009)		(0.009)
substitutions		1.140***	1.665***	1.092***	1.672***	1.104***
		(0.014)	(0.024)	(0.014)	(0.025)	(0.014)
delay		0.994	1.009	0.997	1.009	1.001
		(0.011)	(0.010)	(0.011)	(0.010)	(0.011)
stockout			0.579***		0.541***	
			(0.008)		(0.008)	
missing*substitutions						0.954***
						(0.014)
missing*experience						0.844***
						(0.013)
substitutions*experience						1.091***
						(0.016)
missing*substitutions*experience						0.913***
						(0.017)
40						

(Cont'd on following page)

Table 3.11, continued

	(1)	(2)	(3)	(4)	(5)	(6)
purchase_val	3.935***			5.589***	5.113***	5.602***
	(0.067)			(0.144)	(0.111)	(0.145)
experience	0.833***	1.225***	1.214***	0.812***	0.800***	0.831***
	(0.009)	(0.014)	(0.012)	(0.010)	(0.009)	(0.011)
missing		0.966***		0.890***		0.878***
		(0.013)		(0.012)		(0.012)
substitutions		1.423***	1.530***	1.331***	1.523***	1.350***
		(0.018)	(0.024)	(0.018)	(0.024)	(0.018)
delay		1.128***	1.127***	1.136***	1.132***	1.136***
		(0.012)	(0.011)	(0.013)	(0.011)	(0.013)
stockout			0.860***		0.787***	
			(0.013)		(0.012)	
missing*substitutions						0.981
						(0.014)
missing*experience						0.933***
						(0.015)
substitutions*experience						1.202***
						(0.019)
missing*substitutions*experience						0.954***
						(0.017)
Observations	125,487	73,059	99,258	73,059	99,258	73,059
Pseudo R2	0.034	0.022	0.019	0.057	0.053	0.059

Based on the significant values of the mean and parameter estimates, Store 5 experiences the highest store utility (α Store5 = 1.198), relative to the base store 5, p $\leq$ 0.1), but also has the largest decrease in intrinsic utility if there are missing items (stockouts) (α Store5 = 0.680, p $\leq$ .01). We further observe that the intrinsic utility becomes strongest for Store 5 as the purchase value increases (α Store5 = 3.240, p $\leq$ .01). Lastly, the triple interaction term of missing items (stockouts), substitutions, and customer experience displays a negative utility (coefficient < 1) across all 3 stores. Overall, we can say that across all 3 stores, experienced users are more likely increase their intrinsic utility if substitutions are offered (for stockout items), and have strong intrinsic utility as their purchase value increases.

Overall, the results across all 3 cities indicate some similar findings across all 3 stores

- the most prominent one being that although service quality issues (such as missing items) would strongly decrease the intrinsic utility towards customers primary online store brand, their utility is likely to increase if the customers are more experienced in online grocery shopping, and they are offered substitutions for the stocked-out items. Lastly, we conduct a robustness check to include user demographics of age and gender (estimates for all stores across the 3 cities are presented in Appendix Table C.20).

3.7 Discussion

We motivate our work owing to the lack of substantial research on the drivers of customers' store brand transitions, and customer churn in the online grocery sector. Online grocery retail is a rapidly growing industry, which is expected to reach US$ 800 billion by the end of 2024. However, the industry faces multiple challenges which threaten its survival, with one of the primary concerns being customer churn. In our dataset, for example, assuming a 2-period timeline for churn (i.e., 2-month time for a single purchase), we observe a 56% churn of customers, which is extremely high. In this paper, we therefore investigate the primary drivers of online store switching and platform-exit. Our research was further motivated by relative absence of work on the effects of online service quality issues (such as out-of-stocks) on customers' transition from their primary store of affiliation. We focus on the main effects of service quality parameters such as delays, stockouts (missing items), and substitutions on customer primary store transitions. We further examine the effect of the value of purchases on store choice and store transitions. Next, we investigate the effect of customers' online shopping experience on store-switching, and platform-exit, respectively, when they encounter stockouts and substitutions in the focal time period. Lastly, we examine whether the factors that lead to a store-switch could be utilized to examine customers' choice of specific store brands.

Our findings reveal that service quality parameters such as stockouts and substitutions play a significant role in customers' decision to switch their 'primary' online store-brand, or exit the platform's service. We further observe that customers' online shopping experience has contradictory effects on store-switching, and platform-exit, respectively. Specifically, when experienced customers encounter stockouts (or substitutions), they are less-likely to switch their preferred store-brand. However, experienced customers are more likely to exit the platform after experiencing stockouts in the focal period. This finding indicates that experienced customers exhibit inertia towards their preferred store-brand, but are receptive (i.e., exhibit low stickiness) towards the online platform, when service quality issues arise. Lastly, we also find that customers' probability to exit the platform's service is higher when a customer has had a recent 'primary' store transition. Next, our results from the MNL store choice models across three cities with largest presence of our online platform reveal that although service quality issues (such as missing items) would strongly decrease the intrinsic utility towards customers primary online store brand, their utility is likely to increase if the customers are more experienced in online grocery shopping, and they are offered substitutions for the out-of-stock items.

From a managerial point of view, our results provide useful insights for both online aggregator platforms and retail stores enrolled with online platforms. In particular, we suggest retail stores (enrolled with online platforms) to reduce price promotions and discounts, and instead focus on providing suitable alternatives (substitutions) to products that are likely to be stocked-out, to retain their 'loyal' customers. We further suggest the online grocery platforms to adopt techniques through which picking efficiency (i.e, picking time and quality of picking) could be improved. Improved picking could lead to lesser delays, and higher substitutions in case of stockouts, which in turn, would help retain customers.

Conclusión

En conclusión, esta tesis investiga el rendimiento y la supervivencia de las plataformas de comestibles en línea mediante la utilización de tres lentes separadas - el trabajador gig, la tienda al por menor, y el cliente. Comenzamos examinando los factores de comportamiento específicos de los trabajadores autónomos que podrían afectar al rendimiento, en términos de productividad y calidad del servicio, de las plataformas de comestibles en línea. A continuación, utilizamos esta información para asignar mejor los trabajadores disponibles a las tareas, mediante la incorporación de características específicas tanto del trabajador como de la tarea. Instamos a los gestores de plataformas en línea a utilizar técnicas similares a la hora de asignar tareas, en lugar de un mecanismo de asignación aleatoria.

A continuación, pasamos a los factores externos que podrían afectar al rendimiento de la plataforma. Argumentamos que, en un sector con márgenes operativos estrechos y una configuración altamente competitiva, las plataformas harían mejor en recoger pedidos de tiendas minoristas, en lugar de poseer almacenes propios. Desarrollamos un algoritmo de programación de pedidos que asigna los pedidos no urgentes a las horas ”menos ocupadas” de las tiendas, caracterizadas por un tráfico reducido. Nuestros resultados demuestran que la programación de pedidos reduciría efectivamente el tiempo de entrega, manteniendo un nivel similar de calidad de servicio. Para el resto de pedidos, sugerimos técnicas que eviten las colas, como el uso de cajas automáticas. Por último, también sugerimos la adopción de un mecanismo independiente de recogida y entrega para reducir los retrasos y lograr plazos de entrega más rápidos.

Por último, abordamos los factores que podrían llevar a los clientes a cambiar su tienda online preferida o a abandonar el servicio de la plataforma (churn). En concreto, observamos que los clientes experimentados son más propensos a cambiar de tienda tras sufrir roturas de stock y falta de artículos en el periodo de tiempo focal. Hacemos varias sugerencias a los comercios minoristas y a las plataformas, en particular, para mejorar la utilidad de los clientes. Nuestros resultados revelan que el método más eficaz es evitar las roturas de stock y los retrasos en el servicio. En caso de que el desabastecimiento sea inevitable, la plataforma (y la tienda online) deberían garantizar la disponibilidad de sustitutos adecuados, que deberían ofrecerse a los clientes sin coste adicional. Además, insistimos en que es menos probable que los descuentos y las promociones tengan un efecto duradero en la fidelización de los clientes, y sólo deben utilizarse como último recurso si no se puede ofrecer ningún artículo de sustitución.

Así pues, hemos cerrado el círculo. Nos planteamos la siguiente pregunta: Si tenemos que retener a los clientes de nuestra plataforma, que probablemente abandonen debido a problemas de calidad del servicio, como roturas de stock y retrasos, ¿qué hacemos? Y volvemos la vista a los dos capítulos anteriores para encontrar las respuestas.

Conclusion

In conclusion, this thesis investigates the performance and survival of online grocery platforms by utilizing three separate lenses - the gig worker, the retail store, and the customer. We begin with examining the behavioral factors specific to gig-workers that could affect performance, in terms of productivity, and service quality, of online grocery platforms. We then utilize this information to better allocate the available workers to the tasks, by incorporating both worker- and task-specific characteristics. We urge managers of online platforms to utilize similar techniques while allocating tasks, instead of a random allocation mechanism.

Next, we move to the external factors that could affect the platform's performance. We argue that in an industry with thin operating margins and a highly competitive setup, platforms would be better-off picking orders from retail stores, instead of owning warehouses of their own. We develop an order scheduling algorithm that schedules non-urgent, ambient, grocery orders to 'less-busy' store hours, characterised by low store traffic. Our findings demonstrate that order scheduling would indeed reduce turnaround time, while maintaining a similar level of service quality. For the remaining orders, we suggest queue avoidance techniques such as utilizing self-checkouts. Lastly, we also suggest adopting a separate picker-deliverer mechanism to reduce delays and achieve faster turnarounds.

Finally, we address those factors that could lead to customers' switching their preferred online store, or exiting the platform's service (churn). In particular, we find that experienced customers are more likely to churn after encountering stockouts and missing items in the focal time-period. We make several suggestions to retail stores and platforms, in particular, to improve the utility of customers. Our findings reveal that the most effective method is to avoid in-store stockouts and service delays. In case a stockout is inevitable, the platform (and the online store) should ensure the availability of appropriate substitutes, which should then be offered to the customers at no extra charge. We further stress that providing discounts and promotions is less-likely to have a long-lasting effect on customer retention, and should only be utilized as a last resort if no item substitutions could be offered.

We have therefore come to a full circle. We ask ourselves the question: If we need to retain our platform's customers, who are likely to churn due to service quality issues, such as stockouts and delays, what do we do? And, we look back to our previous two chapters, to find the answers right there!

Bibliography

Aastrup, J. and Kotzab, H. (2009). Analyzing out-of-stock in independent grocery stores: an empirical study. *International Journal of Retail & Distribution Management*, 37(9):765–789.

Abadie, A. and Imbens, G. W. (2006). Large sample properties of matching estimators for average treatment effects. *Econometrica*, 74:235–267.

Abadie, A. and Imbens, G. W. (2011). Bias-corrected matching estimators for average treatment effects. *Journal of Business and Economic Statistics*, 29:1–11.

Agrawal, A., Horton, J. J., Lacetera, N., and Lyons, E. (2015). *Digitization and the contract labor market: A research agenda*, pages 219–250. University of Chicago Press.

Allon, G. (2021). The case against self-checkouts.

Allon, G., Cohen, M. C., and Sinchaisri, W. P. (2023). The impact of behavioral and economic drivers on gig economy workers. *Manufacturing & Service Operations Management*.

Anderson, E., Chen, C., Israeli, A., and Simester, D. (2024). Canary categories. *Journal of Marketing Research*, 0(0):1–19.

Anderson, E. T., Fitzsimons, G. J., and Simester, D. (2006). Measuring and mitigating the costs of stockouts. *Management Science*, 52(11):1751–1763.

Anesbury, Z., Nenycz-Thiel, M., Dawes, J., and Kennedy, R. (2016). How do shoppers behave online? an observational study of online grocery shopping. *Journal of Consumer Behaviour*, 15(3):261–270.

Austin, P. C. (2011). Optimal caliper widths for propensity-score matching when estimating differences in means and differences in proportions in observational studies. *Pharmaceutical Statistics*, 10(2):150–161.

Avgerinos, E. and Gokpinar, B. (2017). Team familiarity and productivity in cardiac surgery operations: The effect of dispersion, bottlenecks, and task complexity. *Manufacturing & Service Operations Management*, 19(1):19–35.

Balov, N. (2016). Bayesian hierarchical models in stata. In *2016 Stata Conference*, number 30. Stata Users Group.

Bansal, H. S., Taylor, S. F., and James, Y. S. (2005). “migrating” to new service providers: Toward a unifying framework of consumers’ switching behaviors. *Journal of the Academy of Marketing Science*, 33(1):96–115.

Batt, R. J. and 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.

Bavafa, H. and Jónasson, J. O. (2021). The variance learning curve. *Management Science*, 67(5):3104–3116.

Belavina, E., Girotra, K., and Kabra, A. (2017). Online grocery retail: Revenue models and environmental impact. *Management Science*, 63(6):1781–1799.

Bell, D. R., Gallino, S., and Moreno, A. (2020). Customer supercharging in experience-centric channels. *Management Science*, 66(9):4096–4107.

Bell, D. R., Ho, T. H., and Tang, C. S. (1998). Determining where to shop: Fixed and variable costs of shopping. *Journal of Marketing Research*, XXXV:352–369.

Bell, D. R. and Lattin, J. M. (1998). Shopping behavior and consumer preference for store price format: Why “large basket” shoppers prefer edlp. *Marketing Science*, 17(1):66–88.

Benjaafar, S., Ding, J.-Y., Kong, G., and Taylor, T. (2022). Labor welfare in on-demand service platforms. *Manufacturing & Service Operations Management*, 24(1):110–124.

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

Boyer, K. K. and 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.

Bozer, Y. A. and Aldarondo, F. J. (2018). A simulation-based comparison of two goods-to-person order picking systems in an online retail setting. *International Journal of Production Research*, 56(11):3838–3858.

Briesch, R. A., Dillon, W. R., and Fox, E. J. (2013). Category positioning and store choice: The role of destination categories. *Marketing Science*, 32(3):488–509.

Campo, K., Gijsbrechts, E., and Nisol, P. (2000). Towards understanding consumer response to stockouts. *Journal of Retailing*, 76(2):219–242.

Chan, T. Y., Li, J., and Pierce, L. (2014). Learning from peers: Knowledge transfer and sales force productivity growth. *Marketing Science*, 33(4):463–484.

Chen, D. and Horton, J. (2016). Research note—are online labor markets spot markets for tasks? a field experiment on the behavioral response to wage cuts. *Information Systems Research*, 27(2):403–423.

Chen, M. K., Chevalier, J. A., Rossi, P. E., and Oehlsen, E. (2019). The value of flexible work: Evidence from uber drivers. *Journal of Political Economy*, 127(6):2735–2794.

Cheng, M. (2021a). Ultra-fast delivery services are ditching the gig-worker model. Accessed on May 15, 2023.

Cheng, M. (2021b). Ultra-fast delivery services are ditching the gig-worker model.

Chintagunta, P. K., Chu, J., and Cebollada, J. (2012). Quantifying transaction costs in online/off-line grocery channel choice. *Marketing Science*, 31(1):96–114.

Chintagunta, P. K., Jain, D. C., and Vilcassim, N. J. (1991). Investigating heterogeneity in brand preferences in logit models for panel data. *Journal of Marketing Research*, 28(4):417–428.

Chou, X., Loske, D., Klumpp, M., Gambardella, L. M., and Montemanni, R. (2021a). *In-store Picking Strategies for Online Orders in Grocery Retail Logistics*, pages 181–189. Springer International Publishing, Cham.

Chou, X., Pietri, N. O., Loske, D., Klumpp, M., and Montemanni, R. (2021b). Optimization strategies for in-store order picking in omnichannel retailing. In Dolgui, A., Bernard, A., Lemoine, D., von Cieminski, G., and Romero, D., editors, *Advances in Production Management Systems. Artificial Intelligence for Sustainable and Resilient Production Systems*, pages 603–611, Cham. Springer International Publishing.

Corsten, D. and Gruen, T. (2003). Desperately seeking shelf availability: an examination of the extent, the causes, and the effort to address retail out-of-stock. *International Journal of Retail & Distribution Management*, 31(12):605–617.

Cousins, C. R. and Tang, N. (2016). Working time and work and family conflict in the netherlands, sweden and the uk. *Work, Employment and Society*, 18(3):531–549.

Dai, H., Milkman, K. L., Hofmann, D. A., and Staats, B. R. (2015). The impact of time at work and time off from work on rule compliance: the case of hand hygiene in health care. *J Appl Psychol*, 100(3):846–62.

Dai, H., Swaminathan, J. M., and Xu, Y. (2022). Leveraging the experience: Exploration and exploitation in gig worker learning process. *Working Paper*.

Dobson, G., Pinker, E., and Van Horn, R. L. (2009). Division of labor in medical office practices. *Manufacturing & Service Operations Management*, 11(3):525–537.

ECREurope (2003). *Optimal Shelf Availability, Increasing Shopper Satisfaction at the Moment of Truth*. ECR Europe.

Espinosa, J. A., Slaughter, S. A., Kraut, R. E., and Herbsleb, J. D. (2007). Familiarity, complexity, and team performance in geographically distributed software development. *Organization Science*, 18(4):613–630.

Ezhil Kumar, M., Sharma, D. P., and Tapar, A. V. (2021). Out-of-stock justifications and consumers' behavioral outcomes– exploring the role of product type and sales level information in out-of-stock situations. *Journal of Retailing and Consumer Services*, 60.

Feldstein, M. S. (1967). Specification of the labour input in the aggregate production function. *The Review of Economic Studies*, 34(4):375–386.

Fickenscher, L. and Wayt, T. (2022). Grocery app gorillas drops 10-minute delivery pledge, adds store pick-up option.

Fox, M. P., Lash, T. L., and Bodnar, L. M. (2020). Common misconceptions about validation studies. *International journal of epidemiology*, 49(4):1392–1396.

Gallino, S. and Moreno, A. (2014). Integration of online and offline channels in retail: The impact of sharing reliable inventory availability information. *Management Science*, 60(6):1434–1451.

Gelman, A., Carlin, J. B., Stern, H. S., Vehtari, A., and Rubin, D. B. (2014). *Bayesian data analysis*. Chapman and Hall/CRC, 3 edition.

Glaeser, C. K., Fisher, M., and Su, X. (2019). Optimal retail location: Empirical methodology and application to practice. *Manufacturing & Service Operations Management*, 21(1):86–102.

Glovo (2021). Algorithms. Accessed on July 18, 2023.

Golden, L. (2012). The effects of working time on productivity and firm performance: a research synthesis paper. *Conditions of Work and Employment Series*, 33.

Gruen, T., Corsten, D., and Bharadwaj, S. (2002). *Retail Out of Stocks: A Worldwide Examination of Extent, Causes and Consumer Responses*. Grocery Manufacturers of America.

Guda, H. and Subramanian, U. (2019). Your uber is arriving: Managing on-demand workers through surge pricing, forecast communication, and worker incentives. *Management Science*.

Guha, R. (2023). First-come, first-serve? gig worker performance and task allocation in an on-demand platform. In *6th European International Conference on Industrial Engineering and Operations Management*. IEOM Society.

Guha, R. and Corsten, D. (2023). The role of within-day learning on gig workers' performance and task allocation: Evidence from an on-demand platform. *SSRN Working Paper*.

Hall, J. V., Horton, J. J., and Knoepfle, D. T. (2021). Pricing in designed markets: The case of ride-sharing. *Working Paper*.

Hansen, K. and Singh, V. (2008). Research note—are store-brand buyers store loyal? an empirical investigation. *Management Science*, 54(10):1828–1834.

He, S., Qiu, L., and Cheng, X. (2021). Surge pricing and short-term wage elasticity of labor supply in real-time ridesharing markets. *Management and Information Systems Quarterly*, 46(1):193–227.

Heckman, J. J. (1979). Sample selection bias as a specification error. *Econometrica*, 47(1):153–161.

Helling, B. (2023). How to get more batches on instacart. Accessed on July 18, 2023.

Hirano, K., Imbens, G., and Ridder, G. (2003). Efficient estimation of average treatment effects using the estimated propensity score. *Econometrica*, 71(4):1161–1189.

Hoang, D. and Breugelmans, E. (2023). “sorry, the product you ordered is out of stock”: Effects of substitution policy in online grocery retailing. *Journal of Retailing*, 99(1):26–45.

Hu, M. M., Yang, S., and Xu, D. Y. (2019). Understanding the social learning effect in contagious switching behavior. *Management Science*, 65(10):4771–4794.

Huckman, R. S. and Pisano, G. P. (2006). The firm specificity of individual performance: Evidence from cardiac surgery. *Management Science*, 52(4):473–488.

Huckman, R. S. and Staats, B. R. (2011). Fluid tasks and fluid teams: The impact of diversity in experience and team familiarity on team performance. *Manufacturing & Service Operations Management*, 13(3):310–328.

Hult, G. T. M., Sharma, P. N., Morgeson, F. V., and Zhang, Y. (2019). Antecedents and consequences of customer satisfaction: Do they differ across online and offline purchases? *Journal of Retailing*, 95(1):10–23.

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

Instacart (2021). When do grocery stores restock? a simple guide for shoppers.

Kc, D. S. (2014). Does multitasking improve performance? evidence from the emergency department. *Manufacturing & Service Operations Management*, 16(2):168–183.

Kc, D. S. (2020). Worker productivity in operations management. *Foundations and Trends® in Technology, Information and Operations Management*, 13(3):151–249.

Kc, D. S. and Staats, B. R. (2012). Accumulating a portfolio of experience: The effect of focal and related experience on surgeon performance. *Manufacturing & Service Operations Management*, 14(4):618–633.

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

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

Kc, D. S. and Terwiesch, C. (2012). An econometric analysis of patient flows in the cardiac intensive care unit. *Manufacturing & Service Operations Management*, 14(1):50–65.

Keaveney, S. M. and Parthasarathy, M. (2001). Customer switching behavior in online services: An exploratory study of the role of selected attitudinal, behavioral, and demographic factors. *Journal of the Academy of Marketing Science*, 29(4):374–390.

Kelliher, C. and Anderson, D. (2010). Doing more with less? flexible working practices and the intensification of work. *Human Relations*, 63(1):83–106.

Kesavan, S., Lambert, S. J., Williams, J. C., and Pendem, P. K. (2022). Doing well by doing good: Improving retail store performance with responsible scheduling practices at the gap, inc. *Management Science*.

Kilbridge, M. and Wester, L. (1966). An economic model for the division of labor. *Management Science*, 12(6):B–255–B–269.

Knight, B., Mitrofanov, D., and Netessine, S. (2023). The impact of ai technology on the productivity of gig economy workers. *Working Paper*.

Krantz, P. (2022). Delivery apps disappear in europe as funding runs out.

Leszczyc, P. T. L. P., Sinha, A., and Timmermans, H. J. P. (2000). Consumer store choice dynamics: An analysis of the competitive market structure for grocery stores. *Journal of Retailing*, 76(3):323–345.

Leszczyc, P. T. L. P. and Timmermans, H. J. P. (1997). Store-switching behavior. *Marketing Letters*, 8(2):193–204.

Lin, W. and Wooldridge, J. M. (2019). *Testing and Correcting for Endogeneity in Nonlinear Unobserved Effects Models*, pages 21–43.

Littek, W. (2001). Division of labor. In Smelser, N. J. and Baltes, P. B., editors, *International Encyclopedia of the Social & Behavioral Sciences*, pages 8220–8226. Pergamon, Oxford.

Lomas, N. (2022). Gopuff to pull out of spain as q-commerce belt tightening continues.

Lu, S. F. and Lu, L. X. (2017). Do mandatory overtime laws improve quality? staffing decisions and operational flexibility of nursing homes. *Management Science*, 63(11):3566–3585.

Lu, Y., Musalem, A., Olivares, M., and Schilkrut, A. (2013). Measuring the effect of queues on customer purchases. *Management Science*, 59(8):1743–1763.

Lupien, S. J., Maheu, F., Tu, M., Fiocco, A., and Schramek, T. E. (2007). The effects of stress and stress hormones on human cognition: Implications for the field of brain and cognition. *Brain and Cognition*, 65(3):209–237.

Melis, K., Campo, K., Breugelmans, E., and Lamey, L. (2015). The impact of the multi-channel retail mix on online store choice: Does online experience matter? *Journal of Retailing*, 91(2):272–288.

Mohamadi, N., Atan, Z., Transchel, S., and Fransoo, J. C. (2023). Impact of delivery radius on the profitability of ultrafast grocery retailers.

Moreno, A. and Gamze, Y. (2020). Migros turkey: Scaling online operations (a).

Moreno, A. and Terwiesch, C. (2015). Pricing and production flexibility: An empirical analysis of the u.s. automotive industry. *Manufacturing & Service Operations Management*, 17(4):428–444.

Musalem, A., Olivares, M., Bradlow, E. T., Terwiesch, C., and Corsten, D. (2010). Structural estimation of the effect of out-of-stocks. *Management Science*, 56(7):1180–1197.

Narayanan, S., Balasubramanian, S., and Swaminathan, J. M. (2009). A matter of balance: Specialization, task variety, and individual learning in a software maintenance environment. *Management Science*, 55(11):1861–1876.

Neves-Moreira, F. and Amorim, P. (2022). Playing hide and seek: tackling in-store picking operations while improving customer experience. *Working Paper*.

O'Brien, S. A. (2022). Two ultra-fast delivery startups shut down in one week.

Paul, S. (2022). Assignment routing optimization for swiggy instamart delivery — part i. Accessed on July 18, 2023.

Pencavel, J. (2015). The productivity of working hours. *The Economic Journal*, 125(589):2052–2076.

Perdikaki, O., Kesavan, S., and Swaminathan, J. M. (2012). Effect of traffic on sales and conversion rates of retail stores. *Manufacturing & Service Operations Management*, 14(1):145–162.

Puligadda, S., Ross, W. T., Chen, J., and Howlett, E. (2012). When loyalties clash purchase behavior when a preferred brand is stocked out: The tradeoff between brand and store loyalty. *Journal of Retailing and Consumer Services*, 19(6):570–577.

Quinn, R. W. (2005). Flow in knowledge work: High performance experience in the design of national security technology. *Administrative Science Quarterly*, 50(4):610–641.

Ramdas, K., Saleh, K., Stern, S., and Liu, H. (2018). Variety and experience: Learning and forgetting in the use of surgical devices. *Management Science*, 64(6):2590–2608.

Raveendran, M., Puranam, P., and Warglien, M. (2022). Division of labor through self-selection. *Organization Science*, 33(2):810–830.

Reagans, R., Argote, L., and Brooks, D. (2005). Individual experience and experience working together: Predicting learning rates from knowing who knows what and knowing how to work together. *Management Science*, 51(6):869–881.

Rhee, H. and Bell, D. R. (2002). The inter-store mobility of supermarket shoppers. *Journal of Retailing*, 78:225–237.

Richards, T. J. and Liaukonytė, J. (2022). Switching cost and store choice. *American Journal of Agricultural Economics*, 105(1):195–218.

Seghezzi, A., Siragusa, C., and Mangiaracina, R. (2022). Enhancing in-store picking for e-grocery: An empirical-based model. *International Journal of Physical Distribution & Logistics Management*, 52(4):301–323.

Semykina, A. and Wooldridge, J. M. (2010). Estimating panel data models in the presence of endogeneity and selection. *Journal of Econometrics*, 157(2):375–380.

Semykina, A. and Wooldridge, J. M. (2013). Estimation of dynamic panel data models with sample selection. *Journal of Applied Econometrics*, 28(1):47–61.

Semykina, A. and Wooldridge, J. M. (2017). Binary response panel data models with sample selection and self-selection. *Journal of Applied Econometrics*, 33(2):179–197.

Sharma, T. (2020). Delivery task allocation: Empowering the delivery performance of delivery agents. Accessed on July 18, 2023.

Sheldon, M. (2016). Income targeting and the ridesharing market. *Working Paper*.

Singh, R. and Rosengren, S. (2020). Why do online grocery shoppers switch? an empirical investigation of drivers of switching in online grocery. *Journal of Retailing and Consumer Services*, 53.

Sloot, L. M. and Verhoef, P. C. (2008). The impact of brand delisting on store switching and brand switching intentions. *Journal of Retailing*, 84(3):281–296.

Sloot, L. M., Verhoef, P. C., and Franses, P. H. (2005). The impact of brand equity and the hedonic level of products on consumer stock-out reactions. *Journal of Retailing*, 81(1):15–34.

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

Statista (2024). Grocery delivery. Accessed on April 15, 2024.

Stock, J. and Yogo, M. (2005). *Testing for Weak Instruments in Linear IV Regression*, pages 80–108. Cambridge University Press, New York.

Su, L., Jiang, Y., Chen, Z., and DEWALL, C. N. (2016). Social exclusion and consumer switching behavior: A control restoration mechanism. *Journal of Consumer Research*, Forthcoming.

Tang, C., Bell, D., and Ho, T.-H. (2001). Store choice and shopping behavior: How price format works. *California Management Review*, 43(2):56–74.

Taylor, T. A. (2018). On-demand service platforms. *Manufacturing & Service Operations Management*, 20(4):704–720.

Thakral, N. and Tô, L. T. (2021). Daily labor supply and adaptive reference points. *American Economic Review*, 111(8):2417–2443.

Ton, Z. and Raman, A. (2010). The effect of product variety and inventory levels on retail store sales: A longitudinal study. *Production and Operations Management*, 19(5):546–560.

Van Lin, A. and Gijsbrechts, E. (2014). Shopper loyalty to whom? chain versus outlet loyalty in the context of store acquisitions. *Journal of Marketing Research*, 51(3):352–370.

Vyt, D., Jara, M., Mevel, O., Morvan, T., and Morvan, N. (2022). The impact of convenience in a click and collect retail setting: A consumer-based approach. *International Journal of Production Economics*, 248:108491.

Wang, K., Gao, Y., Liu, Y., and Nurul Habib, K. (2023). Exploring the choice between in-store versus online grocery shopping through an application of semi-compensatory independent availability logit (scial) model with latent variables. *Journal of Retailing and Consumer Services*, 71.

Wiebach, N. and Hildebrandt, L. (2012). Explaining customers' switching patterns to brand delisting. *Journal of Retailing and Consumer Services*, 19(1):1–10.

Wieland, T. (2021). A micro-econometric store choice model incorporating multi- and omni-channel shopping: The case of furniture retailing in germany. *Geographical Analysis*, 55(1):3–30.

Wooldridge, J. M. (2010). *Econometric Analysis of Cross Section and Panel Data*. The MIT Press, Cambridge, Massachusetts.

Wooldridge, J. M. (2014). Control function methods in applied econometrics. *The Journal of Human Resources*, 50(2):420–445.

Xin, L. and Xie, Y. (2023). The benefits of delay to online decision-making. In *Consortium for Operations Excellence in Retailing*.

Xu, Y., Lu, B., Ghose, A., Dai, H., and Zhou, W. (2023). The interplay of earnings, ratings, and penalties on sharing platforms: An empirical investigation. *Management Science*.

Appendix A

Appendix-Chapter1

Tables

Table A.1: Descriptive Statistics: Choice Equation

Variable	Description	Mean	SD	Min	Max
worked	Whether worker i worked (1) or not (0) in slot t	0.28	0.45	0.00	1.00
avgcomp_last	Avg. compensation received by worker i on the previous work day	9.59	4.29	5.50	29.30
OSF	Same-day orders so far by worker i	1.97	2.22	0.00	25.00
CSF	Same-day compensation so far of worker i	19.19	25.51	0.00	362.90
precipprob	Hourly precipitation probability in city, c	8.43	27.79	0.00	100.00
humidity	Hourly humidity in city, c	68.71	19.96	7.17	100.00
demand	Demand during the same city and slot	33.88	36.60	0.00	208.00
supply	Supply of workers in the same city and slot	30.83	33.43	0.00	190.00
worktime_lag	Worker i's working time on same day as day d in the previous week	2.69	2.64	0.00	13.00
hours_weeklag	Worker i's total working time for the week before the week of day d	16.77	13.31	0.00	79.00
new_worker	1 if the worker is new, 0 otherwise	0.44	0.50	0.00	1.00

Table A.2: Two-stage Heckman: Choice Equation

	(1) worked
avgcomp_last	0.071***
	(0.001)
OSF	0.328***
	(0.004)
CSF	-0.033***
	(0.000)
precipprob	-0.001***
	(0.000)
humidity	-0.002***
	(0.000)
demand	0.026***
	(0.000)
supply	-0.024***
	(0.000)
worktime_lag	0.037***
	(0.000)
hours_weeklag	0.015***
	(0.000)
new_worker	0.000
	(0.001)
resid1	-0.344***
	(0.004)
resid2	0.034***
	(0.000)
constant	-1.678***
	(0.008)
Observations	5957844
Pseudo R2	0.056

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table A.3: Correlation Matrix

Variables	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)
pickingtime	1.000													
delay	0.187	1.000												
	(0.000)													
substituted_when_reqd	0.378	0.102	1.000											
	(0.000)	(0.000)												
OSF	-0.142	0.054	0.029	1.000										
	(0.000)	(0.000)	(0.000)											
OSF_other	-0.051	0.077	0.049	0.495	1.000									
	(0.000)	(0.000)	(0.000)	(0.000)										
avgcomp_last	-0.035	0.062	0.150	0.034	0.011	1.000								
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)									
avgstockout_other	0.032	0.056	0.087	0.109	0.244	0.073	1.000							
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)								
experience	-0.215	-0.084	-0.046	0.214	0.045	-0.109	-0.172	1.000						
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)							
advancetime	0.172	-0.512	0.015	-0.041	-0.076	-0.015	-0.048	0.031	1.000					
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)						
num_item	0.660	0.122	0.363	-0.052	-0.022	-0.035	0.017	-0.025	0.115	1.000				
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)					
num_substitution	0.378	0.102	1.000	0.029	0.049	0.150	0.087	-0.046	0.015	0.363	1.000			
	(0.000)	(0.000)	(1.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)				
call_duration	0.069	0.026	0.042	0.002	0.010	-0.007	0.010	-0.015	0.015	0.023	0.042	1.000		
	(0.000)	(0.000)	(0.000)	(0.017)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)			
storefamiliarity	-0.155	-0.074	-0.108	0.171	0.020	-0.185	-0.158	0.779	0.083	0.048	-0.108	-0.016	1.000	
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)		
sqm'000s	0.111	0.042	0.013	-0.007	-0.002	0.000	-0.031	-0.002	0.022	-0.049	0.013	0.003	0.024	1.000
	(0.000)	(0.000)	(0.000)	(0.000)	(0.008)	(0.867)	(0.000)	(0.043)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	

Table A.4: OSF vs Picking Time (discrete measures for batching, complexity, and discretion)

	(1)	(2)	(3)	(4)	(5)	(6)
OSF	-0.065	-0.087*	-0.785***	-0.115*	0.139***	0.119***
	(0.054)	(0.046)	(0.078)	(0.064)	(0.045)	(0.036)
CSF	-0.026***	-0.027***	-0.020***	-0.026***	-0.021**	-0.021**
	(0.008)	(0.008)	(0.007)	(0.008)	(0.008)	(0.009)
experience	-0.644***	-0.644***	-0.899***	-0.606***	-0.624***	-0.617***
	(0.017)	(0.017)	(0.033)	(0.018)	(0.016)	(0.017)
num_item	0.629***	0.629***	0.629***	0.633***	0.610***	0.608***
	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)
num_substitution	0.981***	0.981***	0.981***	0.985***	0.533***	0.499***
	(0.003)	(0.003)	(0.003)	(0.003)	(0.001)	(0.004)
call_duration	1.525***	1.525***	1.526***	1.524***	1.448***	1.423***
	(0.027)	(0.027)	(0.026)	(0.026)	(0.026)	(0.025)
storefamiliarity	-0.931***	-0.932***	-0.930***	-0.986***	-0.889***	-0.893***
	(0.018)	(0.018)	(0.018)	(0.019)	(0.021)	(0.019)
store_size	1.223***	1.223***	1.223***	1.222***	1.225***	1.225***
	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)
precipprob	-0.002***	-0.002***	-0.003***	-0.002***	-0.002***	-0.002***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
humidity	-0.004***	-0.004***	-0.003***	-0.004***	-0.004***	-0.004***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
demand	0.099***	0.099***	0.109***	0.057***	0.098***	0.098***
	(0.005)	(0.005)	(0.005)	(0.006)	(0.005)	(0.005)
supply	-0.097***	-0.097***	-0.107***	-0.052***	-0.096***	-0.096***
	(0.004)	(0.004)	(0.003)	(0.004)	(0.003)	(0.003)
worktime_lag	0.008	0.007	0.016	0.007	0.010	0.010
	(0.016)	(0.016)	(0.015)	(0.016)	(0.016)	(0.015)
hours_weeklag	-0.024***	-0.024***	-0.021***	-0.025***	-0.023***	-0.023***
	(0.004)	(0.004)	(0.004)	(0.004)	(0.003)	(0.003)
IMR	0.644**	0.617*	0.975***	0.663**	0.639**	0.624**
	(0.312)	(0.316)	(0.290)	(0.328)	(0.299)	(0.304)
OSF2		0.003***				
		(0.001)				
OSF_exp			0.116***			
			(0.007)			
OSF_batch				0.198***		
				(0.022)		
batch_pick				0.881***		
				(0.065)		
OSF_complex					-0.194***	
					(0.008)	
med_complexity					3.029***	
					(0.051)	
high_complexity					5.444***	
					(0.075)	
OSF_discretion						-0.159***
						(0.014)
med_discretion						2.754***
						(0.042)
high_discretion						5.422***
						(0.086)
Picker & Time fixed effects	yes	yes	yes	yes	yes	yes

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table A.5: OSF vs Delay (discrete measures for batching, complexity, and discretion)

	(1)	(2)	(3)	(4)	(5)	(6)
OSF	0.184^{***}	0.825^{***}	1.516^{***}	0.205^{***}	-0.020	-0.027
	(0.010)	(0.026)	(0.066)	(0.003)	(0.067)	(0.061)
CSF	0.044^{***}	0.048^{***}	0.033^{***}	0.044^{***}	0.040^{***}	0.040^{***}
	(0.006)	(0.001)	(0.004)	(0.005)	(0.005)	(0.005)
experience	-0.044	-0.031	0.414^{***}	-0.055	-0.046	-0.046
	(0.038)	(0.037)	(0.014)	(0.038)	(0.037)	(0.038)
advancetime	-0.282^{***}	-0.282^{***}	-0.281^{***}	-0.282^{***}	-0.282^{***}	-0.282^{***}
	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)
pickingtime	0.319^{***}	0.321^{***}	0.320^{***}	0.323^{***}	0.320^{***}	0.321^{***}
	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)
deliverytime	0.328^{***}	0.329^{***}	0.328^{***}	0.320^{***}	0.328^{***}	0.328^{***}
	(0.008)	(0.008)	(0.008)	(0.008)	(0.008)	(0.008)
precipprob	0.001	-0.000	0.002	0.001	0.001	0.001
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
humidity	0.006^{***}	0.009^{***}	0.005^{***}	0.006^{***}	0.006^{***}	0.006^{***}
	(0.001)	(0.001)	(0.001)	(0.001)	(0.002)	(0.002)
demand	0.053^{***}	0.071^{***}	0.037^{***}	0.000	0.050^{***}	0.050^{***}
	(0.014)	(0.009)	(0.014)	(0.013)	(0.014)	(0.014)
supply	-0.057^{***}	-0.078^{***}	-0.041^{***}	0.001	-0.054^{***}	-0.054^{***}
	(0.015)	(0.010)	(0.015)	(0.014)	(0.015)	(0.015)
worktime_lag	-0.051^{*}	-0.026	-0.064^{**}	-0.052^{*}	-0.054^{**}	-0.053^{**}
	(0.027)	(0.021)	(0.027)	(0.027)	(0.027)	(0.027)
hours_weeklag	-0.028^{***}	-0.018^{***}	-0.032^{***}	-0.029^{***}	-0.029^{***}	-0.028^{***}
	(0.008)	(0.005)	(0.008)	(0.008)	(0.008)	(0.008)
IMR	-1.659^{**}	-0.875^{*}	-2.209^{***}	-1.615^{**}	-1.753^{**}	-1.733^{**}
	(0.756)	(0.487)	(0.746)	(0.755)	(0.759)	(0.759)
OSF2		-0.098^{***}				
		(0.001)				
OSF_exp			-0.209^{***}			
			(0.012)			
OSF_batch				-0.112^{***}		
				(0.017)		
batch_pick				2.016^{***}		
				(0.084)		
OSF_complex					0.159^{***}	
					(0.045)	
med_complexity					-0.745^{***}	
					(0.104)	
high_complexity					-0.287	
					(0.175)	
OSF_discretion						0.142^{***}
						(0.033)
med_discretion						-0.793^{***}
						(0.048)
high_discretion						-0.693^{***}
						(0.102)
Picker fixed effects	yes	yes	yes	yes	yes	yes
Day fixed effect	yes	yes	yes	yes	yes	yes

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table A.6: OSF vs Substituted_when_reqd (discrete measures for batching, complexity, discretion)

	(1)	(2)	(3)	(4)	(5)	(6)
OSF	0.045***	0.039***	0.041***	0.045***	0.002	0.002
	(0.004)	(0.003)	(0.012)	(0.004)	(0.002)	(0.002)
CSF	-0.000	-0.000	0.002***	-0.000	0.001*	0.001**
	(0.001)	(0.001)	(0.000)	(0.001)	(0.000)	(0.000)
resid3	-0.031***	-0.032***	-0.037***	-0.032***	0.001	-0.004***
	(0.002)	(0.003)	(0.013)	(0.002)	(0.001)	(0.001)
resid4	-0.001*	-0.001*	-0.001	-0.001*	-0.001***	-0.001***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.000)	(0.000)
experience	0.085***	0.085***	0.084***	0.084***	0.043***	0.050***
	(0.001)	(0.001)	(0.004)	(0.001)	(0.001)	(0.001)
num_item	0.038***	0.038***	0.038***	0.038***	0.019***	0.019***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
storefamiliarity	-0.149***	-0.149***	-0.149***	-0.148***	-0.050***	-0.061***
	(0.000)	(0.000)	(0.012)	(0.000)	(0.000)	(0.001)
precipprob	-0.000***	-0.000***	-0.000***	-0.000***	-0.000***	-0.000***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
humidity	0.001***	0.001***	0.001***	0.001***	0.000***	0.000***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
demand	0.002***	0.002***	0.002***	0.003***	0.001***	0.001***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
supply	-0.001***	-0.001**	-0.001***	-0.003***	0.000	0.000
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
worktime_lag	0.000	-0.000	0.000	0.000	0.002***	0.001***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
hours_weeklag	0.001***	0.001***	0.001***	0.001***	0.001***	0.001***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
IMR	0.046**	0.041*	0.047**	0.047***	0.075***	0.057***
	(0.018)	(0.022)	(0.022)	(0.017)	(0.010)	(0.018)
IMR2	1.797***	1.797***	1.797***	1.798***	1.787***	1.578***
	(0.011)	(0.011)	(0.011)	(0.011)	(0.006)	(0.009)
OSF*OSF		0.001***				
		(0.000)				
resid3*resid3		0.001***				
		(0.000)				
OSF*experience			0.001			
			(0.001)			
resid3*experience			0.001			
			(0.002)			
batch_pick				-0.054***		
				(0.003)		
batch_pick*OSF				-0.001**		
				(0.001)		
batch_pick*resid3				0.008***		
				(0.002)		
med_complexity					1.216***	
					(0.003)	
high_complexity					1.979***	
					(0.001)	
med_complexity*OSF					0.002***	
					(0.000)	
high_complexity*OSF					-0.001	
					(0.001)	
med_complexity*resid3					-0.002*	
					(0.001)	
high_complexity*resid3					-0.001***	
					(0.000)	
med_discretion						1.419***
						(0.001)
high_discretion						2.278***
						(0.001)
med_discretion*OSF						0.001*
						(0.001)
high_discretion*OSF						-0.000
						(0.001)
med_discretion*resid3						0.003***
						(0.001)
high_discretion*resid3						0.003***
						(0.000)
Picker & Day fixed effects	yes	yes	yes	yes	yes	yes

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table A.7: Two-stage Heckman Selection - HSF vs Picking Time

	(1)	(2)	(3)	(4)	(5)	(6)
HSF	-0.169***	-0.139***	-0.927***	-0.200***	-0.119**	-0.153***
	(0.051)	(0.034)	(0.061)	(0.062)	(0.047)	(0.048)
CSF	-0.019***	-0.019***	-0.015**	-0.020***	-0.013*	-0.018***
	(0.007)	(0.007)	(0.006)	(0.008)	(0.007)	(0.007)
experience	-0.649***	-0.649***	-0.895***	-0.601***	-0.629***	-0.656***
	(0.017)	(0.017)	(0.023)	(0.020)	(0.018)	(0.018)
num_item	0.629***	0.629***	0.629***	0.634***	0.598***	0.628***
	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)
num_substitution	0.982***	0.982***	0.982***	0.987***	0.123***	1.501***
	(0.003)	(0.003)	(0.003)	(0.003)	(0.014)	(0.006)
call_duration	1.526***	1.526***	1.527***	1.523***	1.392***	1.113***
	(0.027)	(0.027)	(0.026)	(0.027)	(0.023)	(0.026)
storefamiliarity	-0.932***	-0.931***	-0.931***	-0.999***	-0.848***	-0.939***
	(0.017)	(0.017)	(0.018)	(0.020)	(0.021)	(0.020)
store_size	1.223***	1.223***	1.222***	1.221***	1.226***	1.211***
	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)
precipprob	-0.002***	-0.002***	-0.003***	-0.002***	-0.002***	-0.002***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
humidity	-0.004***	-0.004***	-0.003***	-0.004***	-0.005***	-0.004***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
demand	0.103***	0.103***	0.110***	0.053***	0.096***	0.102***
	(0.005)	(0.004)	(0.005)	(0.005)	(0.005)	(0.004)
supply	-0.101***	-0.101***	-0.108***	-0.046***	-0.094***	-0.100***
	(0.003)	(0.003)	(0.004)	(0.003)	(0.004)	(0.003)
worktime_lag	0.014	0.015	0.019	0.014	0.013	0.014
	(0.014)	(0.014)	(0.015)	(0.014)	(0.016)	(0.013)
hours_weeklag	-0.021***	-0.021***	-0.019***	-0.021***	-0.021***	-0.021***
	(0.003)	(0.003)	(0.004)	(0.003)	(0.003)	(0.003)
IMR	0.855***	0.885***	1.083***	0.972***	0.731**	0.862***
	(0.256)	(0.233)	(0.294)	(0.261)	(0.287)	(0.256)
HSF2		-0.005*				
		(0.003)				
HSF_exp			0.123***			
			(0.003)			
HSF_batch				0.091***		
				(0.009)		
num_batched				0.615***		
				(0.027)		
HSF_complex					-0.065***	
					(0.002)	
ccomplexity					1.009***	
					(0.003)	
HSF_discretion						-0.017***
						(0.005)
cdiscretion						-1.010***
						(0.006)
Picker fixed effects	yes	yes	yes	yes	yes	yes
Day fixed effect	yes	yes	yes	yes	yes	yes

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table A.8: Two-stage Heckman Selection - HSF vs Delay

	(1)	(2)	(3)	(4)	(5)	(6)
HSF	0.202***	0.961***	1.594***	0.219***	0.095***	0.194***
	(0.006)	(0.035)	(0.084)	(0.008)	(0.029)	(0.011)
CSF	0.045***	0.048***	0.037***	0.045***	0.040***	0.044***
	(0.005)	(0.002)	(0.004)	(0.004)	(0.004)	(0.005)
experience	-0.040	-0.031	0.411***	-0.050	-0.039	-0.041
	(0.037)	(0.037)	(0.009)	(0.038)	(0.037)	(0.037)
advancetime	-0.282***	-0.282***	-0.281***	-0.282***	-0.281***	-0.282***
	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)
pickingtime	0.319***	0.321***	0.320***	0.322***	0.318***	0.321***
	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.002)
deliverytime	0.328***	0.329***	0.328***	0.320***	0.328***	0.328***
	(0.008)	(0.008)	(0.008)	(0.008)	(0.008)	(0.008)
precipprob	0.001	-0.000	0.001	0.001	0.001	0.001
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
humidity	0.006***	0.009***	0.005***	0.005***	0.005***	0.006***
	(0.002)	(0.001)	(0.001)	(0.001)	(0.002)	(0.002)
demand	0.057***	0.074***	0.044***	0.001	0.055***	0.056***
	(0.014)	(0.010)	(0.013)	(0.013)	(0.013)	(0.014)
supply	-0.061***	-0.081***	-0.048***	0.000	-0.058***	-0.060***
	(0.015)	(0.012)	(0.015)	(0.015)	(0.015)	(0.015)
worktime_lag	-0.045*	-0.020	-0.055**	-0.046*	-0.048*	-0.046*
	(0.027)	(0.023)	(0.026)	(0.027)	(0.027)	(0.027)
hours_weeklag	-0.025***	-0.016***	-0.028***	-0.026***	-0.026***	-0.025***
	(0.008)	(0.005)	(0.007)	(0.008)	(0.008)	(0.008)
IMR	-1.441*	-0.669	-1.853***	-1.374*	-1.557**	-1.473*
	(0.760)	(0.559)	(0.699)	(0.759)	(0.739)	(0.756)
HSF2		-0.134***				
		(0.002)				
HSF_exp			-0.226***			
			(0.015)			
HSF_batch				-0.056***		
				(0.011)		
num_batched				1.026***		
				(0.032)		
HSF_complex					0.051***	
					(0.012)	
ccomplexity					-0.089***	
					(0.021)	
HSF_discretion						0.019**
						(0.008)
cdiscretion						-0.125***
						(0.024)
Picker fixed effects	yes	yes	yes	yes	yes	yes
Day fixed effect	yes	yes	yes	yes	yes	yes

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table A.9: Two-stage Heckman Selection - HSF vs Substituted when requested

	(1)	(2)	(3)	(4)	(5)	(6)
HSF	0.046***	0.036***	0.042***	0.046***	0.004*	0.038***
	(0.005)	(0.004)	(0.014)	(0.006)	(0.002)	(0.004)
CSF	0.000	0.000	0.000	0.000	0.000	0.000
	(0.001)	(0.001)	(0.001)	(0.001)	(0.000)	(0.001)
resid3	-0.034***	-0.036***	-0.042***	-0.035***	0.004***	-0.030***
	(0.003)	(0.003)	(0.013)	(0.002)	(0.000)	(0.002)
resid4	-0.001*	-0.001*	-0.001*	-0.001*	-0.001***	-0.001*
	(0.001)	(0.001)	(0.001)	(0.001)	(0.000)	(0.000)
experience	0.086***	0.086***	0.084***	0.084***	0.019***	0.056***
	(0.001)	(0.001)	(0.004)	(0.001)	(0.002)	(0.002)
IMR	0.028	0.021	0.029	0.029	-0.016	0.053**
	(0.020)	(0.025)	(0.023)	(0.019)	(0.022)	(0.022)
IMR2	1.797***	1.797***	1.797***	1.798***	1.690***	1.752***
	(0.011)	(0.011)	(0.011)	(0.011)	(0.005)	(0.011)
HSF*HSF		0.002***				
		(0.000)				
resid3*resid3		0.001***				
		(0.000)				
HSF*experience			0.001			
			(0.001)			
resid3*experience			0.002			
			(0.002)			
num_batched				-0.025***		
				(0.002)		
num_batched*HSF				-0.000		
				(0.000)		
num_batched*resid3				0.004***		
				(0.001)		
ccomplexity					0.203***	
					(0.001)	
ccomplexity*HSF					-0.001***	
					(0.000)	
ccomplexity*resid3					0.000	
					(0.002)	
cdiscretion						0.158***
						(0.002)
cdiscretion*HSF						-0.001***
						(0.000)
cdiscretion*resid3						0.001
						(0.000)
Picker fixed effects	yes	yes	yes	yes	yes	yes
Day fixed effect	yes	yes	yes	yes	yes	yes
Controls	Yes	Yes	Yes	Yes	Yes	Yes

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table A.10: Two-stage Selection - OSF vs Picking Time (Alternate IV)

	(1)	(2)	(3)	(4)	(5)	(6)
OSF	-0.074	-0.074	-0.320	-0.123	-0.279	-0.077
	(0.131)	(0.131)	(0.369)	(0.182)	(0.398)	(0.159)
CSF	-0.025*	-0.074	-0.022	-0.025	0.003	-0.025
	(0.014)	(0.131)	(0.019)	(0.018)	(0.038)	(0.018)
experience	-0.645***	-0.074	-0.725***	-0.596***	-0.629***	-0.652***
	(0.020)	(0.131)	(0.090)	(0.024)	(0.005)	(0.022)
num_item	0.629***	-0.074	0.629***	0.634***	0.597***	0.628***
	(0.004)	(0.131)	(0.004)	(0.004)	(0.004)	(0.004)
num_substitution	0.981***	-0.074	0.981***	0.986***	0.123***	1.500***
	(0.003)	(0.131)	(0.003)	(0.004)	(0.015)	(0.006)
call_duration	1.525***	-0.074	1.526***	1.522***	1.394***	1.113***
	(0.026)	(0.131)	(0.025)	(0.025)	(0.027)	(0.025)
storefamiliarity	-0.931***	-0.074	-0.931***	-0.999***	-0.843***	-0.939***
	(0.019)	(0.131)	(0.020)	(0.022)	(0.013)	(0.023)
store_size	1.223***	-0.074	1.223***	1.221***	1.226***	1.211***
	(0.007)	(0.131)	(0.007)	(0.007)	(0.005)	(0.007)
precipprob	-0.002***	-0.074	-0.002***	-0.002***	-0.002**	-0.002***
	(0.001)	(0.131)	(0.001)	(0.001)	(0.001)	(0.001)
humidity	-0.003***	-0.074	-0.003***	-0.003***	-0.004***	-0.003***
	(0.001)	(0.131)	(0.001)	(0.001)	(0.001)	(0.001)
demand	0.092***	-0.074	0.094***	0.038***	0.087***	0.090***
	(0.013)	(0.131)	(0.012)	(0.014)	(0.006)	(0.012)
supply	-0.091***	-0.074	-0.093***	-0.032***	-0.087***	-0.089***
	(0.011)	(0.131)	(0.010)	(0.012)	(0.005)	(0.010)
worktime_lag	-0.002	-0.074	-0.001	-0.009	-0.000***	-0.005
	(0.025)	(0.131)	(0.026)	(0.027)	(0.000)	(0.024)
hours_weeklag	-0.028***	-0.074	-0.028***	-0.032***	-0.027***	-0.027***
	(0.008)	(0.131)	(0.009)	(0.010)	(0.004)	(0.009)
IMR	0.250	-0.074	0.283	0.058	0.136	0.147
	(0.755)	(0.131)	(0.784)	(0.873)	(0.449)	(0.755)
OSF2		0.192***				
		(0.031)				
OSF_exp			0.036			
			(0.031)			
OSF_batch				0.089***		
				(0.009)		
num_batched				0.583***		
				(0.030)		
OSF_complex					-0.059***	
					(0.011)	
ccomplexity					1.011***	
					(0.040)	
OSF_discretion						-0.002
						(0.018)
cdiscretion						-1.043***
						(0.038)
Picker fixed effects	yes	yes	yes	yes	yes	yes
Day fixed effect	yes	yes	yes	yes	yes	yes

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table A.11: Two-stage Selection - OSF vs Delay (Alternate IV)

	(1)	(2)	(3)	(4)	(5)	(6)
OSF	0.727	1.598	2.524	0.717	0.604	0.719
	(1.637)	(7.881)	(3.454)	(1.645)	(1.604)	(1.658)
CSF	-0.008	-0.029	-0.032	-0.007	-0.016	-0.011
	(0.144)	(0.523)	(0.227)	(0.146)	(0.158)	(0.151)
experience	-0.036	-0.045	0.547	-0.046	-0.038	-0.038
	(0.068)	(0.257)	(0.437)	(0.069)	(0.066)	(0.067)
advancetime	-0.282***	-0.284***	-0.281***	-0.283***	-0.282***	-0.282***
	(0.002)	(0.010)	(0.003)	(0.002)	(0.003)	(0.003)
pickingtime	0.319***	0.326***	0.320***	0.322***	0.319***	0.321***
	(0.002)	(0.022)	(0.001)	(0.003)	(0.000)	(0.002)
deliverytime	0.328***	0.325***	0.327***	0.319***	0.328***	0.328***
	(0.010)	(0.005)	(0.011)	(0.009)	(0.010)	(0.010)
precipprob	0.000	-0.003	0.001	0.001	0.001	0.000
	(0.004)	(0.017)	(0.006)	(0.004)	(0.004)	(0.004)
humidity	0.007	0.007	0.006	0.007	0.006	0.007
	(0.005)	(0.028)	(0.008)	(0.005)	(0.006)	(0.005)
demand	0.058	0.210	0.046	0.001	0.056	0.057
	(0.060)	(0.500)	(0.078)	(0.061)	(0.060)	(0.060)
supply	-0.059	-0.215	-0.046	0.003	-0.056	-0.058
	(0.052)	(0.467)	(0.065)	(0.054)	(0.051)	(0.052)
worktime_lag	-0.043	-0.060	-0.049	-0.047	-0.046	-0.045
	(0.089)	(0.638)	(0.123)	(0.089)	(0.091)	(0.091)
hours_weeklag	-0.024	0.020	-0.025	-0.026	-0.024	-0.025
	(0.039)	(0.260)	(0.055)	(0.039)	(0.041)	(0.040)
IMR	-1.307	0.937	-1.527	-1.362	-1.359	-1.352
	(3.790)	(23.485)	(5.300)	(3.800)	(3.904)	(3.865)
OSF2		-0.096				
		(0.296)				
OSF_exp			-0.265*			
			(0.160)			
OSF_batch				-0.019		
				(0.019)		
num_batched				0.939***		
				(0.070)		
OSF_complex					0.066	
					(0.061)	
ccomplexity					-0.144	
					(0.152)	
OSF_discretion						0.043
						(0.054)
cdiscretion						-0.189
						(0.142)
Picker fixed effects	yes	yes	yes	yes	yes	yes
Day fixed effect	yes	yes	yes	yes	yes	yes

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table A.12: Two-stage Selection - OSF vs Substituted when requested (Alternate IV)

	(1)	(2)	(3)	(4)	(5)	(6)
OSF	0.180	0.175	0.171	0.180	0.014	0.152
	(0.310)	(0.319)	(0.299)	(0.309)	(0.018)	(0.253)
CSF	-0.013	-0.013	-0.013	-0.013	-0.001	-0.011
	(0.028)	(0.029)	(0.028)	(0.028)	(0.001)	(0.023)
resid3	-0.163	-0.165	-0.159	-0.164	-0.007	-0.139
	(0.310)	(0.319)	(0.295)	(0.310)	(0.019)	(0.254)
resid4	0.011	0.011	0.011	0.011	0.000	0.010
	(0.028)	(0.029)	(0.028)	(0.028)	(0.001)	(0.023)
experience	0.089***	0.089***	0.086***	0.088***	0.019***	0.059***
	(0.008)	(0.008)	(0.005)	(0.008)	(0.001)	(0.006)
IMR	-0.027	-0.044	-0.029	-0.026	0.011	0.038
	(0.682)	(0.701)	(0.669)	(0.681)	(0.059)	(0.577)
IMR2	1.794***	1.794***	1.794***	1.794***	1.690***	1.749***
	(0.004)	(0.004)	(0.004)	(0.004)	(0.005)	(0.005)
OSF*OSF		0.001***				
		(0.000)				
resid3*resid3		-0.000				
		(0.000)				
OSF*experience			0.001			
			(0.001)			
resid3*experience			-0.000			
			(0.002)			
num_batched				-0.028***		
				(0.002)		
num_batched*OSF				0.001**		
				(0.000)		
batch_pick*resid3				0.002*		
				(0.001)		
ccomplexity					0.203***	
					(0.001)	
ccomplexity*OSF					-0.001***	
					(0.000)	
ccomplexity*resid3					-0.001***	
					(0.000)	
cdiscretion						0.158***
						(0.002)
cdiscretion*OSF						-0.001**
						(0.000)
cdiscretion*resid3						0.000
						(0.000)
Picker fixed effects	yes	yes	yes	yes	yes	yes
Day fixed effect	yes	yes	yes	yes	yes	yes
Controls	yes	yes	yes	yes	yes	yes

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table A.13: Subsample of workers with above-average experience

	(1)	(2)	(3)
OSF	-0.053	0.707***	0.022***
	(0.075)	(0.106)	(0.005)
OSF2	0.002	-0.087***	0.001***
	(0.006)	(0.021)	(0.000)
CSF	-0.026***	0.049***	0.001***
	(0.002)	(0.008)	(0.000)
experience	-0.569***	-0.077***	0.086***
	(0.020)	(0.023)	(0.001)
num_item	0.607***		0.039***
	(0.002)		(0.000)
num_substitution	0.974***		
	(0.035)		
call_duration	1.560***		
	(0.021)		
storefamiliarity	-0.897***		-0.146***
	(0.073)		(0.001)
store_size	1.202***		
	(0.014)		
precipprob	-0.003***	0.001	0.000
	(0.000)	(0.001)	(0.000)
humidity	-0.002	0.011***	0.001***
	(0.001)	(0.000)	(0.000)
demand	0.100***	0.054***	-0.002
	(0.012)	(0.012)	(0.003)
supply	-0.097***	-0.062***	0.003
	(0.014)	(0.010)	(0.003)
worktime_lag	0.017***	-0.054***	-0.006**
	(0.004)	(0.007)	(0.003)
hours_weeklag	-0.019***	-0.029***	-0.002
	(0.003)	(0.001)	(0.001)
IMR	0.981***	-1.973***	-0.194
	(0.360)	(0.187)	(0.122)
advancetime		-0.271***	
		(0.004)	
pickingtime		0.310***	
		(0.005)	
deliverytime		0.316***	
		(0.005)	
resid3			-0.015***
			(0.003)
resid3*resid3			0.001***
			(0.000)
resid4			-0.002***
			(0.000)
IMR2			1.813***
			(0.015)
Picker fixed effects	yes	yes	yes
Day fixed effect	yes	yes	yes

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table A.14: Subsample of workers with average+1SD experience

	(1)	(2)	(3)
OSF	0.115***	0.443***	-0.028***
	(0.007)	(0.095)	(0.006)
OSF2	0.002	-0.076***	0.001***
	(0.003)	(0.005)	(0.000)
CSF	-0.038***	0.063***	0.006***
	(0.003)	(0.013)	(0.001)
experience	-0.412***	-0.149**	0.082***
	(0.154)	(0.072)	(0.003)
num_item	0.583***		0.039***
	(0.006)		(0.000)
num_substitution	0.961***		
	(0.047)		
call_duration	1.586***		
	(0.042)		
storefamiliarity	-0.765***		-0.139***
	(0.061)		(0.001)
store_size	1.198***		
	(0.021)		
precipprob	-0.004***	0.003***	0.001***
	(0.001)	(0.000)	(0.000)
humidity	-0.004	0.015***	0.002***
	(0.002)	(0.001)	(0.000)
demand	0.132***	0.007	-0.011***
	(0.026)	(0.012)	(0.002)
supply	-0.127***	-0.020**	0.011***
	(0.028)	(0.009)	(0.002)
worktime_lag	0.045***	-0.083***	-0.015***
	(0.005)	(0.016)	(0.000)
hours_weeklag	-0.007	-0.041***	-0.005***
	(0.007)	(0.002)	(0.000)
IMR	2.327***	-3.795***	-0.609***
	(0.442)	(0.443)	(0.036)
advancetime		-0.266***	
		(0.002)	
pickingtime		0.305***	
		(0.007)	
deliverytime		0.309***	
		(0.001)	
resid3			0.035***
			(0.010)
resid3*resid3			0.001***
			(0.000)
resid4			-0.006***
			(0.001)
IMR2			1.819***
			(0.006)
Picker fixed effects	yes	yes	yes
Day fixed effect	yes	yes	yes

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table A.15: Subsample of workers with OSF in the top 25%

	(1)	(2)	(3)
OSF	0.144**	0.503***	-0.020***
	(0.063)	(0.079)	(0.005)
OSF2	-0.003***	-0.084***	0.001***
	(0.000)	(0.010)	(0.000)
CSF	-0.035***	0.063***	0.004***
	(0.006)	(0.016)	(0.001)
experience	-0.591***	-0.058	0.082***
	(0.042)	(0.063)	(0.005)
num_item	0.588***		0.039***
	(0.002)		(0.000)
num_substitution	0.929***		
	(0.007)		
call_duration	1.565***		
	(0.001)		
storefamiliarity	-0.772***		-0.141***
	(0.020)		(0.002)
store_size	1.187***		
	(0.029)		
precipprob	-0.005***	0.004***	0.001***
	(0.001)	(0.001)	(0.000)
humidity	-0.007*	0.016***	0.002***
	(0.003)	(0.005)	(0.000)
demand	0.153***	-0.012	-0.017***
	(0.023)	(0.016)	(0.001)
supply	-0.148***	-0.001	0.016***
	(0.020)	(0.011)	(0.001)
worktime_lag	0.066*	-0.106***	-0.019***
	(0.040)	(0.031)	(0.005)
hours_weeklag	-0.007*	-0.040***	-0.005***
	(0.004)	(0.010)	(0.001)
IMR	3.411**	-4.879***	-0.842***
	(1.402)	(1.336)	(0.140)
advancetime		-0.274***	
		(0.006)	
pickingtime		0.314***	
		(0.005)	
deliverytime		0.318***	
		(0.006)	
resid3			0.027***
			(0.009)
resid3*resid3			0.001***
			(0.000)
resid4			-0.005***
			(0.001)
IMR2			1.808***
			(0.006)
Picker fixed effects	yes	yes	yes
Day fixed effect	yes	yes	yes

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table A.16: Subsample of workers after removing bottom 10% inactive workers

	(1)	(2)	(3)
OSF	-0.087***	0.824***	0.039***
	(0.021)	(0.026)	(0.007)
OSF2	0.003	-0.098***	0.001***
	(0.003)	(0.003)	(0.000)
CSF	-0.027***	0.048***	-0.000
	(0.005)	(0.005)	(0.000)
experience	-0.645***	-0.031*	0.085***
	(0.011)	(0.018)	(0.000)
num_item	0.629***		0.038***
	(0.002)		(0.000)
num_substitution	0.981***		
	(0.013)		
call_duration	1.525***		
	(0.056)		
storefamiliarity	-0.931***		-0.150***
	(0.059)		(0.005)
store_size	1.223***		
	(0.017)		
precipprob	-0.002	-0.000	-0.000
	(0.001)	(0.001)	(0.000)
humidity	-0.004*	0.009***	0.001**
	(0.002)	(0.001)	(0.000)
demand	0.099***	0.072***	0.002
	(0.017)	(0.020)	(0.004)
supply	-0.097***	-0.078***	-0.001
	(0.014)	(0.019)	(0.003)
worktime_lag	0.007	-0.025	-0.000
	(0.032)	(0.030)	(0.006)
hours_weeklag	-0.024	-0.018	0.001
	(0.019)	(0.012)	(0.003)
IMR	0.619	-0.849	0.042
	(1.329)	(0.984)	(0.241)
advancetime		-0.282***	
		(0.006)	
pickingtime		0.321***	
		(0.005)	
deliverytime		0.329***	
		(0.004)	
resid3			-0.032***
			(0.007)
resid3*resid3			0.001***
			(0.000)
resid4			-0.001***
			(0.000)
IMR2			1.797***
			(0.001)
Picker fixed effects	yes	yes	yes
Day fixed effect	yes	yes	yes

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table A.17: 2-week lag worker behaviors

	(1)	(2)	(3)
OSF	-0.117***	0.875***	0.058***
	(0.009)	(0.037)	(0.003)
OSF2	0.003*	-0.098***	0.001***
	(0.002)	(0.000)	(0.000)
CSF	-0.023***	0.044***	-0.002***
	(0.002)	(0.001)	(0.000)
experience	-0.610***	-0.041	0.083***
	(0.015)	(0.031)	(0.000)
num_item	0.629***		0.038***
	(0.004)		(0.000)
num_substitution	0.982***		
	(0.003)		
call_duration	1.526***		
	(0.026)		
storefamiliarity	-0.932***		-0.150***
	(0.017)		(0.000)
store_size	1.223***		
	(0.007)		
precipprob	-0.002***	-0.001	-0.000***
	(0.000)	(0.001)	(0.000)
humidity	-0.003***	0.008***	0.000***
	(0.001)	(0.001)	(0.000)
demand	0.092***	0.084***	0.006***
	(0.002)	(0.007)	(0.001)
supply	-0.090***	-0.090***	-0.005***
	(0.003)	(0.009)	(0.001)
worktime_lag	-0.046***	-0.029	0.011***
	(0.001)	(0.026)	(0.003)
hours_2weeklag	-0.100***	-0.010	0.016***
	(0.002)	(0.022)	(0.002)
IMR	0.145	-0.182	0.300***
	(0.151)	(0.447)	(0.068)
advancetime		-0.282***	
		(0.003)	
pickingtime		0.321***	
		(0.003)	
deliverytime		0.329***	
		(0.008)	
resid3			-0.054***
			(0.002)
resid3*resid3			0.001***
			(0.000)
resid4			0.001***
			(0.000)
IMR2			1.797***
			(0.016)
Picker fixed effects	yes	yes	yes
Day fixed effect	yes	yes	yes

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table A.18: Alternative measure of weather (actual hourly precipitation)

	(1)	(2)	(3)
OSF	0.017	0.776***	0.027***
	(0.009)	(0.000)	(0.001)
OSF2	0.000	-0.090***	0.001***
	(0.002)	(0.000)	(0.000)
CSF	-0.032***	0.045***	0.001
	(0.003)	(0.003)	(0.000)
experience	-0.589***	-0.082**	0.079***
	(0.010)	(0.037)	(0.000)
num_item	0.629***		0.038***
	(0.004)		(0.000)
num_substitution	0.981***		
	(0.003)		
call_duration	1.525***		
	(0.027)		
storefamiliarity	-0.933***		-0.150***
	(0.017)		(0.000)
store_size	1.223***		
	(0.007)		
precip_hourly	-0.066***	0.039***	0.001***
	(0.004)	(0.005)	(0.000)
demand	0.113***	0.077***	0.000
	(0.002)	(0.001)	(0.000)
supply	-0.109***	-0.082***	0.001***
	(0.002)	(0.001)	(0.000)
worktime_lag	0.054***	-0.020***	-0.003***
	(0.003)	(0.007)	(0.001)
hours_weeklag	-0.025***	-0.009***	0.000
	(0.000)	(0.001)	(0.000)
IMR	1.503***	-0.397***	-0.055***
	(0.115)	(0.071)	(0.021)
advancetime		-0.282***	
		(0.003)	
pickingtime		0.321***	
		(0.003)	
deliverytime		0.329***	
		(0.008)	
resid3			-0.025***
			(0.002)
resid3*resid3			0.001***
			(0.000)
resid4			-0.001***
			(0.000)
IMR2			1.797***
			(0.016)
Picker fixed effects	yes	yes	yes
Day fixed effect	yes	yes	yes

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table A.19: Regressions including peakhour controls

	(1)	(2)	(3)
OSF	-0.199***	0.837***	0.024***
	(0.013)	(0.010)	(0.000)
OSF2	0.027***	-0.093***	0.001***
	(0.002)	(0.001)	(0.000)
CSF	-0.026***	0.045***	0.001***
	(0.002)	(0.004)	(0.000)
experience	-0.617***	-0.036	0.084***
	(0.017)	(0.034)	(0.000)
num_item	0.629***		0.038***
	(0.004)		(0.000)
num_substitution	0.984***		
	(0.003)		
call_duration	1.524***		
	(0.027)		
storefamiliarity	-0.933***		-0.149***
	(0.018)		(0.000)
store_size	1.222***		
	(0.007)		
peakhour	1.216***	0.234***	-3.321***
	(0.144)	(0.001)	(0.390)
precipprob	-0.001***	-0.001*	-0.000***
	(0.000)	(0.000)	(0.000)
humidity	-0.001***	0.008***	0.001***
	(0.000)	(0.001)	(0.000)
demand	0.112***	0.083***	0.000
	(0.001)	(0.001)	(0.000)
supply	-0.112***	-0.089***	0.000
	(0.002)	(0.001)	(0.000)
worktime_lag	0.052***	-0.011*	-0.003***
	(0.003)	(0.006)	(0.001)
hours_weeklag	-0.024***	-0.009***	0.000
	(0.000)	(0.001)	(0.000)
IMR	1.545***	-0.164*	-0.040*
	(0.113)	(0.086)	(0.023)
advancetime		-0.282***	
		(0.003)	
pickingtime		0.321***	
		(0.003)	
deliverytime		0.329***	
		(0.008)	
resid3			-0.019***
			(0.001)
resid3*resid3			0.001***
			(0.000)
resid4			-0.002***
			(0.000)
IMR2			1.798***
			(0.011)
Picker fixed effects	yes	yes	yes
Day fixed effect	yes	yes	yes

$^{*}\ p < 0.10$, $^{**}\ p < 0.05$, $^{***}\ p < 0.01$

Table A.20: OSF vs Picking Time: Including OSF2 in models m4 to m6

	batching	complexity	discretion
OSF	-0.133**	-0.050	-0.075*
	(0.056)	(0.043)	(0.043)
OSF2	0.002**	0.002**	0.003***
	(0.001)	(0.001)	(0.001)
CSF	-0.026***	-0.019**	-0.026***
	(0.008)	(0.008)	(0.008)
experience	-0.595***	-0.622***	-0.651***
	(0.019)	(0.018)	(0.018)
IMR	0.648**	0.470	0.653**
	(0.324)	(0.321)	(0.302)
OSF_cbatch	0.094***		
	(0.009)		
num_batched	0.571***		
	(0.031)		
OSF_ccomplex		-0.059***	
		(0.002)	
ccomplexity		1.010***	
		(0.003)	
OSF_cdiscretion			-0.017***
			(0.005)
cdiscretion			-1.006***
			(0.006)
Picker fixed effects	yes	yes	yes
Day fixed effects	yes	yes	yes

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table A.21: OSF vs Delay: Including OSF2 in models m3 and m4

	experience	batching
OSF	1.760***	0.807***
	(0.015)	(0.040)
OSF_exp	-0.156***	
	(0.010)	
OSF2	-0.089***	-0.099***
	(0.001)	(0.001)
CSF	0.040***	0.048***
	(0.000)	(0.000)
experience	0.310***	-0.039
	(0.017)	(0.037)
IMR	-1.353***	-0.866*
	(0.523)	(0.505)
OSF_cbatch		0.042***
		(0.014)
num_batched		1.538***
		(0.119)
Picker fixed effects	yes	yes
Day fixed effects	yes	yes

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table A.22: 1st stage results - Picking time (m1)

	(1) OSF	(2) CSF	(3) pickingtime
OSF_other	0.938***	-9.326***	
	(0.004)	(0.039)	
CSF_other	0.006***	1.712***	
	(0.000)	(0.003)	
OSF			-0.065***
			(0.018)
CSF			-0.026***
			(0.001)
experience	-0.080***	-1.020***	-0.644***
	(0.002)	(0.020)	(0.005)
num_item	-0.005***	-0.051***	0.629***
	(0.000)	(0.001)	(0.001)
num_substitution	0.009***	0.067***	0.981***
	(0.001)	(0.007)	(0.005)
call_duration	0.005	0.017	1.525***
	(0.003)	(0.035)	(0.004)
storefamiliarity	0.028***	0.271***	-0.931***
	(0.001)	(0.016)	(0.001)
store_size	-0.009***	-0.081***	1.223***
	(0.001)	(0.008)	(0.014)
precipprob	0.003***	0.236***	-0.002***
	(0.000)	(0.001)	(0.000)
humidity	0.001***	0.313***	-0.004***
	(0.000)	(0.001)	(0.000)
demand	-0.058***	-4.839***	0.099***
	(0.001)	(0.012)	(0.000)
supply	0.057***	4.488***	-0.097***
	(0.001)	(0.011)	(0.000)
worktime_lag	-0.019***	-6.475***	0.008***
	(0.002)	(0.017)	(0.003)
hours_weeklag	-0.019***	-3.002***	-0.024***
	(0.001)	(0.007)	(0.003)
IMR	-2.722***	-276.064***	0.644***
	(0.055)	(0.603)	(0.169)
Observations	6046793	6046793	1661655
R2			0.589
Picker fixed effects	yes	yes	yes
Day fixed effect	yes	yes	yes

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table A.23: 1st stage results - Delay (m1)

	(1) OSF	(2) CSF	(3) Delay
OSF_other	0.941***	-9.297***	
	(0.004)	(0.039)	
CSF_other	0.005***	1.707***	
	(0.000)	(0.003)	
OSF			0.184***
			(0.023)
CSF			0.044***
			(0.002)
experience	-0.066***	-0.879***	-0.044***
	(0.001)	(0.015)	(0.005)
advancetime	-0.001***	-0.014***	-0.282***
	(0.000)	(0.000)	(0.000)
pickingtime	-0.006***	-0.057***	0.319***
	(0.000)	(0.001)	(0.001)
deliverytime	0.001***	0.004***	0.328***
	(0.000)	(0.001)	(0.001)
precipprob	0.003***	0.235***	0.001**
	(0.000)	(0.001)	(0.001)
humidity	0.001***	0.311***	0.006***
	(0.000)	(0.001)	(0.001)
demand	-0.057***	-4.819***	0.053***
	(0.001)	(0.012)	(0.015)
supply	0.055***	4.469***	-0.057***
	(0.001)	(0.011)	(0.016)
worktime_lag	-0.018***	-6.461***	-0.051***
	(0.002)	(0.017)	(0.006)
hours_weeklag	-0.019***	-2.996***	-0.028***
	(0.001)	(0.007)	(0.005)
IMR	-2.660***	-275.388***	-1.659***
	(0.055)	(0.603)	(0.397)
Observations	6046793	6046793	1661655
R2			0.400
Picker fixed effects	yes	yes	yes
Day fixed effect	yes	yes	yes

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table A.24: 1st stage results - Substituted when requested (m1)

	(1) OSF	(2) CSF
OSF_other	0.932^{***}	-9.443^{***}
	(0.205)	(2.251)
CSF_other	0.006	1.721^{***}
	(0.019)	(0.131)
experience	-0.079^{***}	-1.008^{***}
	(0.004)	(0.028)
num_item	-0.004^{***}	-0.047^{***}
	(0.000)	(0.005)
storefamiliarity	0.026^{***}	0.244^{***}
	(0.005)	(0.067)
precipprob	0.003	0.239^{***}
	(0.002)	(0.046)
humidity	0.001	0.316^{***}
	(0.004)	(0.035)
demand	-0.060	-4.880^{***}
	(0.051)	(0.913)
supply	0.058	4.525^{***}
	(0.047)	(0.863)
worktime_lag	-0.021	-6.522^{***}
	(0.074)	(1.217)
hours_weeklag	-0.020	-3.028^{***}
	(0.031)	(0.568)
IMR	-2.821	-278.315^{***}
	(2.844)	(49.788)
IMR2	0.045^{***}	0.409^{***}
	(0.002)	(0.016)
Observations	6046793	6046793

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Figures

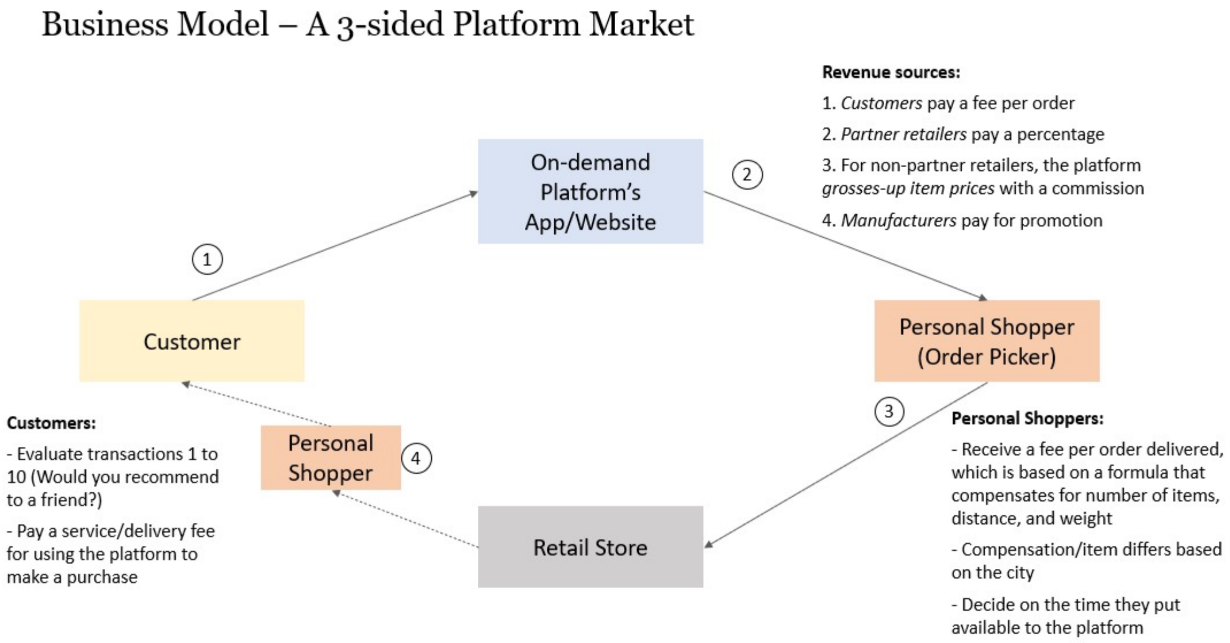
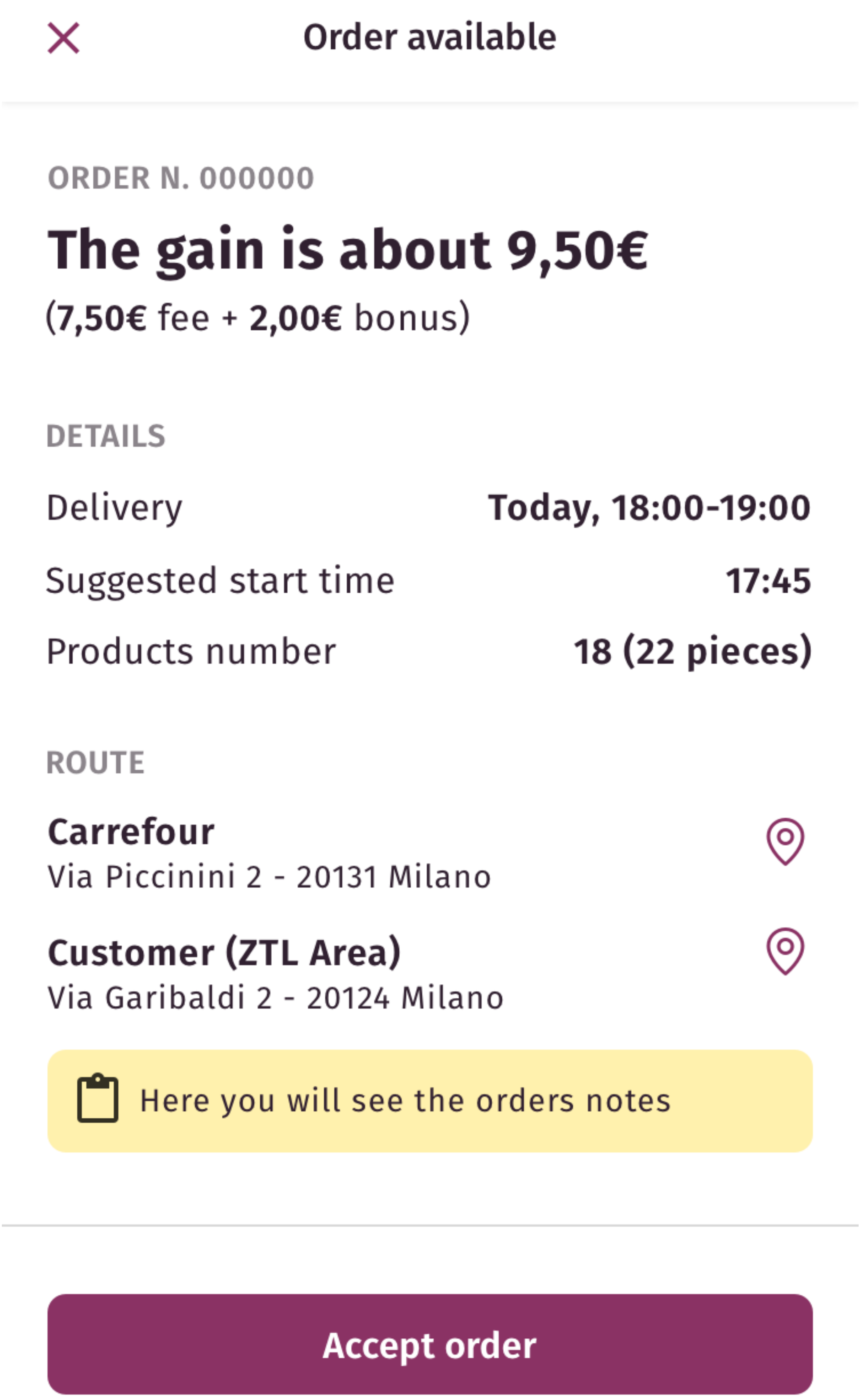


Figure A.1: Business Model of on-demand grocery platforms



ORDER N. 000000

The gain is about 9,50€

(**7,50€** fee + **2,00€** bonus)

DETAILS

Delivery	**Today, 18:00-19:00**
Suggested start time	**17:45**
Products number	**18 (22 pieces)**

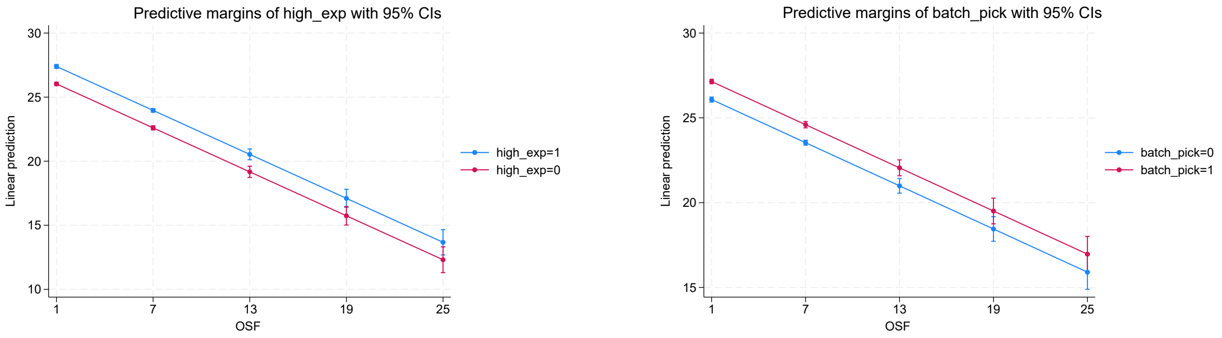
ROUTE

Carrefour
Via Piccinini 2 - 20131 Milano

Customer (ZTL Area)
Via Garibaldi 2 - 20124 Milano

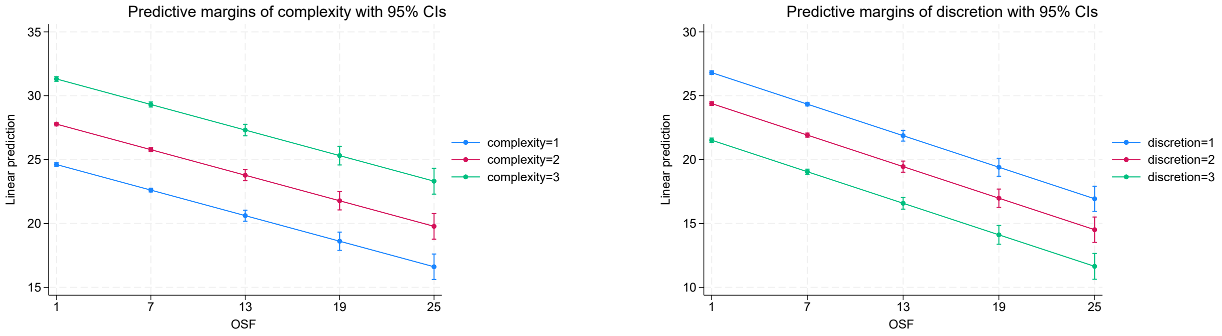
Here you will see the orders notes

Figure A.2: Landing Page of a Personal Shopper



(a) OSF*Experience

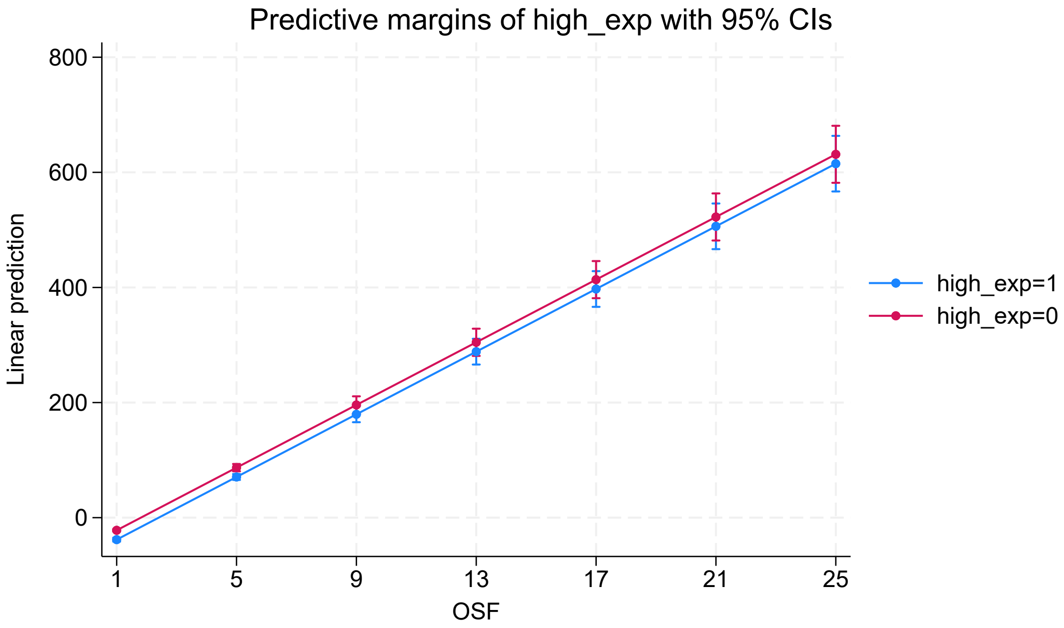
(b) OSF*Batch pick



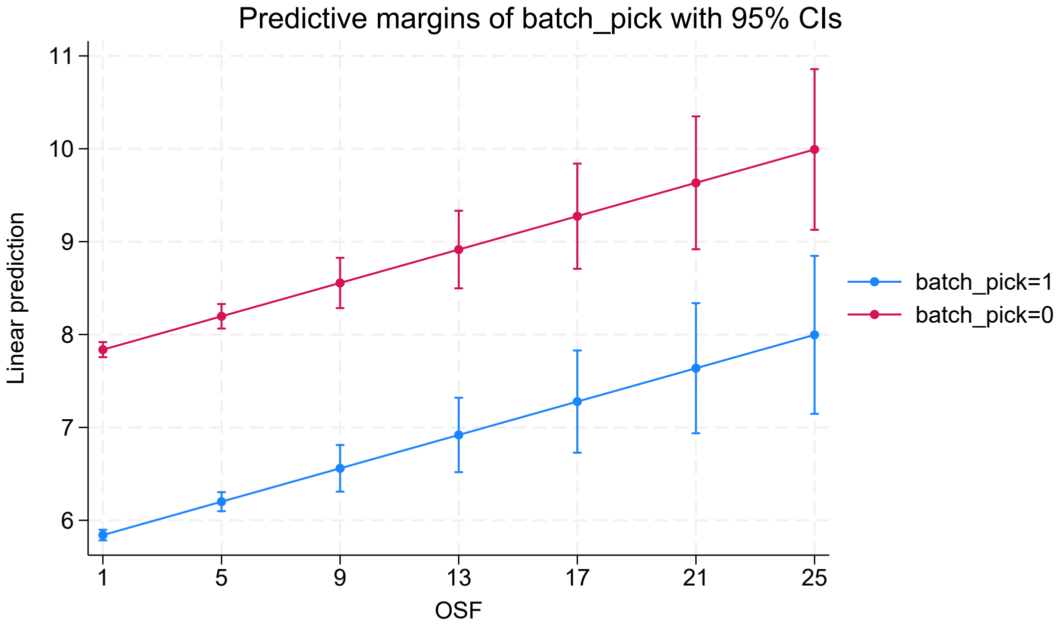
(c) OSF*Complexity

(d) OSF*Discretion

Figure A.3: Interaction effects of OSF on Picking Time

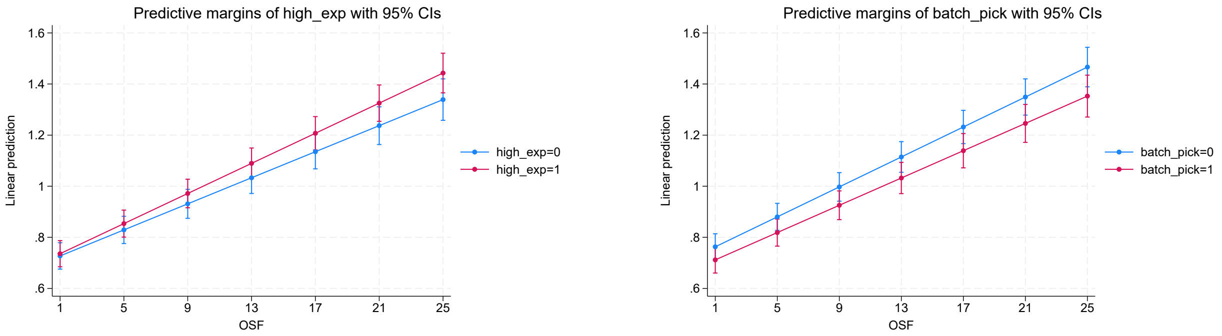

Predictive margins of high_exp with 95% CIs

(a) OSF*Experience


Predictive margins of batch_pick with 95% CIs

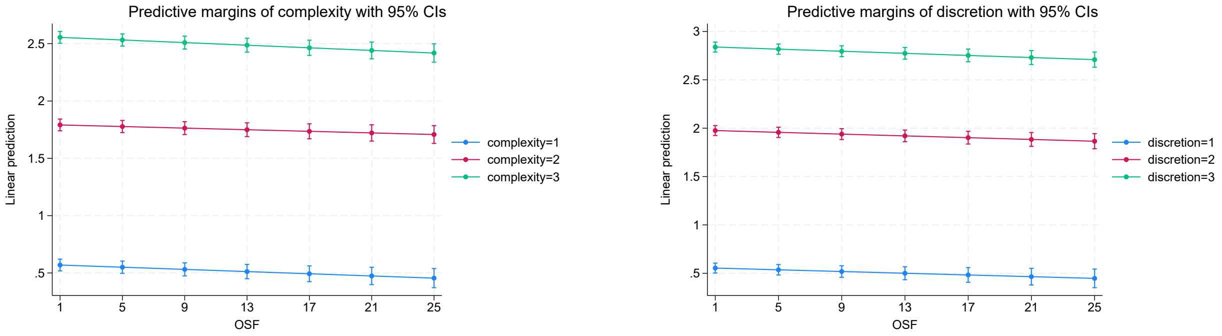
(b) OSF*Batch pick

Figure A.4: Interaction effects of OSF on Delay



(a) OSF*Experience

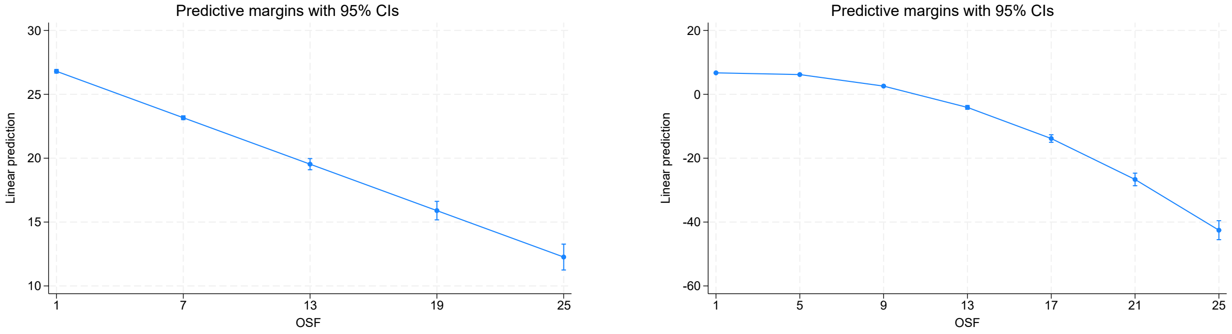
(b) OSF*Batch pick



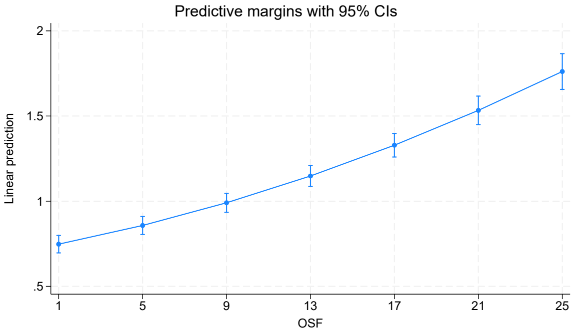
(c) OSF*Complexity

(d) OSF*Discretion

Figure A.5: Interaction effects of OSF on Substituted when requested



(a) OSF2 vs Picking Time (b) OSF2 vs Delay



(c) OSF2 vs Substituted when requested

Figure A.6: Plots of OSF squared on picking time, delay, and substituted when requested

Appendix B

Appendix-Chapter2

Tables

Table B.1: Matching estimations using Nearest Neighbor Matching (NNM)

	(1) pickingtime	(2) num_stockout
ATE		
r1vs0.emptystore_queue_store	-1.683***	-0.077***
	(0.072)	(0.016)
N	2,579,112	2,579,112

$^{+}$ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Models 1 and 2 of Table 10 represents the matching estimation carried out using Nearest Neighbor Matching (NNM) for picking time, and number of stockouts, respectively. The results of the estimations are in line with our main estimation for picking time. However, we observe that the number of stockouts decreases when the orders are picked in a less-busy store.

Models 1 and 2 of Table 11 represent the matching estimation carried out using Inverse Probability Weighting (IPW) for picking time, and number of stockouts, respectively. Models 3 and 4 represent the Regression Adjustment (RA) estimates for picking time, and number of stockouts, respectively. Finally, Models 5 and 6 represent the IPWRA (Inverse

Table B.2: Matching estimations using IPW, RA, and IPWRA

	(1) pickingtime	(2) num_stockout	(3) pickingtime	(4) num_stockout	(5) pickingtime	(6) num_stockout
ATE						
emptystore_queue	-2.031**	1.373***	-2.134***	0.011	-2.136***	0.005
	(0.766)	(0.154)	(0.034)	(0.009)	(0.034)	(0.009)
POmean						
emptystore_queue	26.833***	3.388***	26.888***	3.384***	26.888***	3.384***
	(0.010)	(0.002)	(0.010)	(0.002)	(0.010)	(0.002)
N	2,579,112	2,579,112	2,579,112	2,579,112	2,579,112	2,579,112

$^{+}$ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Probability Weighted Regression Adjustment) estimation, which is a combination of the IPW and RA models, for picking time, and number of stockouts, respectively. The results of the estimations are in line with our main estimation. For models 4 and 6, the increase in number of stockouts is not statistically significant.

Table B.3: PSM - Alternative Models for Robustness

	(1) pickingtime	(2) pickingtime	(3) pickingtime	(4) num_stockout	(5) num_stockout	(6) num_stockout
ATE						
emptystore_queue	-1.852***	-1.888***	-1.877***	0.027**	0.027**	0.030**
	(0.039)	(0.036)	(0.039)	(0.010)	(0.010)	(0.011)
N	2,579,112	2,579,112	2,579,112	2,579,112	2,579,112	2,579,112

$^{+}$ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Models 1 and 4 of Table 12 represent the Propensity Score (PS) matching estimation carried out using a probit model, instead of a logit model, for picking time, and number of stockouts, respectively. Models 2 and 5 represent the PSM estimates for picking time, and number of stockouts, respectively using the number of matches equal to 2, instead of 1 (which is the default). Finally, Models 3 and 6 represent the PSM estimates for picking time, and number of stockouts, respectively, without using a caliper of 0.2 (used in the main estimation). The results of the estimations are in line with our main estimation.

Model 1 of Table 13 represents the average change of +0.17 missing items if all orders were picked from a less-busy store (treatment=1). The result is similar in both sign and

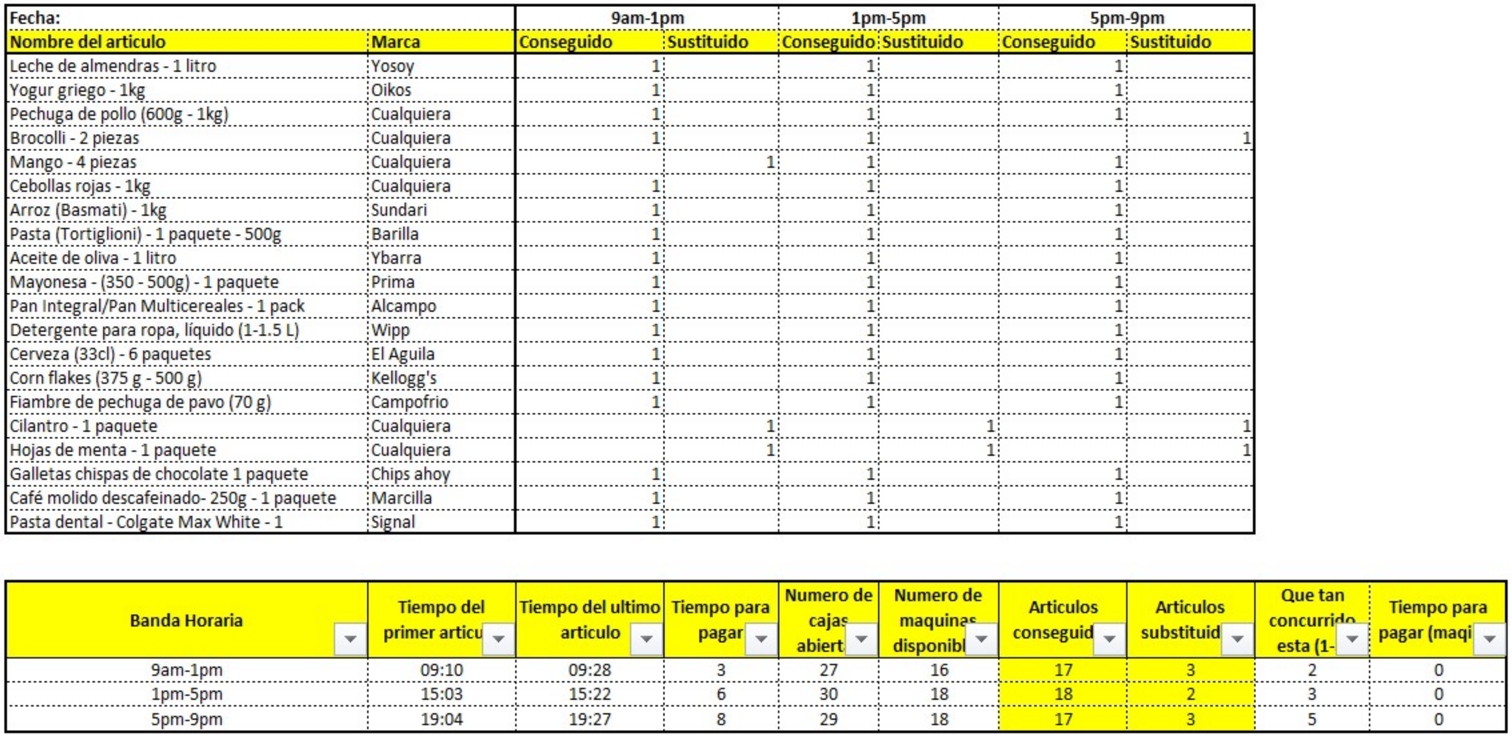
Table B.4: PSM - Alternative DV (No. of missing items)

	(1) num_missing
ATE	
r1vs0.emptystore_queue_store	0.170^{***}
	(0.008)
N	2,579,112

$^{+}$ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

magnitude to the estimations performed using number of stockouts as the dependent variable.

Figures



Fecha:		9am-1pm		1pm-5pm		5pm-9pm	
Nombre del articulo	Marca	Conseguido	Sustituido	Conseguido	Sustituido	Conseguido	Sustituido
Leche de almendras - 1 litro	Yosoy	1		1		1	
Yogur griego - 1kg	Oikos	1		1		1	
Pechuga de pollo (600g - 1kg)	Cualquiera	1		1		1	
Brocolli - 2 piezas	Cualquiera	1		1			1
Mango - 4 piezas	Cualquiera		1	1		1	
Cebollas rojas - 1kg	Cualquiera	1		1		1	
Arroz (Basmati) - 1kg	Sundari	1		1		1	
Pasta (Tortiglioni) - 1 paquete - 500g	Barilla	1		1		1	
Aceite de oliva - 1 litro	Ybarra	1		1		1	
Mayonesa - (350 - 500g) - 1 paquete	Prima	1		1		1	
Pan Integral/Pan Multicereales - 1 pack	Alcampo	1		1		1	
Detergente para ropa, líquido (1-1.5 L)	Wipp	1		1		1	
Cerveza (33cl) - 6 paquetes	El Aguila	1		1		1	
Corn flakes (375 g - 500 g)	Kellogg's	1		1		1	
Fiambre de pechuga de pavo (70 g)	Campofrio	1		1		1	
Cilantro - 1 paquete	Cualquiera		1		1		1
Hojas de menta - 1 paquete	Cualquiera		1		1		1
Galletas chispas de chocolate 1 paquete	Chips ahoy	1		1		1	
Café molido descafeinado- 250g - 1 paquete	Marcilla	1		1		1	
Pasta dental - Colgate Max White - 1	Signal	1		1		1	

Banda Horaria	Tiempo del primer articu	Tiempo del ultimo articulo	Tiempo para pagar	Numero de cajas abiert	Numero de maquinas disponibl	Articulos conseguid	Articulos substituid	Que tan concurrido esta (1-	Tiempo para pagar (maqi
9am-1pm	09:10	09:28	3	27	16	17	3	2	0
1pm-5pm	15:03	15:22	6	30	18	18	2	3	0
5pm-9pm	19:04	19:27	8	29	18	17	3	5	0

Figure B.1: List of items

PARTICIPANT INFORMATION SHEET

You are invited to participate in a research study conducted by IE University.

Project title	Analysis of customer interference on picking operations in retail stores
Principal Investigator and co-investigators	Reeju Guha +34 622236908 reejuguha@student.ie.edu IE Business School Dr Daniel Corsten +41 794550283 Daniel.Corsten@ie.edu IE Business School

1. What is the purpose of this research?

The purpose of the study is to understand how productivity (picking time), and service quality of order pickers (who pick groceries from retail stores) might be affected due to increased store traffic, and does using self-checkout counters, and mobile scan-on-the-go reduce picking time?

2. What is the expected duration of your participation?

The total duration of participation is 4 weeks.

3. What will you need to do?

You will be asked to go to a grocery retail store and 'pick' items from a list provided to you. You are not required to physically pick or pay for any of the items. You need to scan the bar codes of the items using any scanning app on your phone. After the picking of items is complete, you are required to stand in the checkout queue till the person ahead of you pays, which is when you exit the queue, and record that time. You will also be asked to complete an excel sheet consisting of information regarding the number of checkout counters open, the number of items not found, and/or substituted with another brand, and your perception of store traffic (scale of 1-5). You are required to perform this activity three times a day, in different time slots, on each day of the week. The duration of the task would be 10-30 minutes. Each day, you will be asked to send the filled-in excel sheet, along with a screenshot of the list of barcodes.

4. How will your privacy and the confidentiality of your data be protected?

Please see Annex A below.

5. What are the possible discomforts and risks for you?

There are no risks expected during your participation in this study. To reduce any discomforts, you are free to select any time to perform the activity within a given slot.

6. Will there be reimbursement for participation?

Yes. You will receive 105€ at the end of each week, if you complete all excel sheets and submit all required screenshots. The total amount for the participation will be 420€ for the entire duration of four weeks. All earnings are in the form of an account transfer that would be made to you at the end of each week of the study.

7. Can you decide not to participate in this research?

Yes. The decision to participate in this research is entirely voluntary. You can withdraw from the research at any time by informing the principal investigator. If you decide to withdraw in between an ongoing week (for e.g., week 2), all the data collected from you in that week (i.e., week 2) will be discarded, and you will not be reimbursed for that particular week. However, you will receive monetary incentives till the previous week (i.e., till week 1). Please see Annex A for more details on how the data collected from you is treated.

8. Whom should you call if you have any questions?

Please contact for all research-related matters:
Reeju Guha, IE Business School; **Phone**: +34 622 23 69 08; **Email**: reejuguha@student.ie.edu

Figure B.2: Participant Information Sheet (GDPR)

Appendix C

Appendix-Chapter3

Tables

Table C.1: Gibbs Sampling Parameter Estimates (Store-switching and Platform Exit)

	Store-switching	Platform-exit (2-month)
purchase_val	-0.031**	-0.024**
	(0.001)	(0.001)
delay_std	0.008**	-0.006**
	(0.001)	(0.001)
experience	-0.064**	-0.053**
	(0.001)	(0.001)
missing_std	0.054**	0.031**
	(0.001)	(0.001)
substitutions_std	0.054**	0.027**
	(0.001)	(0.001)
constant	0.470**	0.194**
	(0.001)	(0.001)
Observations	435,737	448,537

** $p < 0.05$

Table C.2: MH Sampling Parameter Estimates (Store-switching, Covid19 control)

	(MH-1)	(MH-2)
purchase_val	-0.084**	-0.084**
	(0.003)	(0.002)
delay_std	0.015**	0.015**
	(0.002)	(0.002)
experience	-0.207**	-0.216**
	(0.002)	(0.002)
missing_std	0.135**	0.133**
	(0.002)	(0.002)
substitutions_std	0.136**	0.133**
	(0.002)	(0.002)
missing_std*experience		-0.010**
		(0.003)
substitutions_std*experience		-0.029**
		(0.003)
missing_std*substitutions_std		-0.008**
		(0.002)
missing_std*substitutions_std*experience		0.002
		(0.003)
is_covid	0.069**	0.064**
	(0.005)	(0.005)
constant	-0.093**	-0.094**
	(0.002)	(0.002)
Observations	435,737	435,737

** $p < 0.05$

Table C.3: MH Sampling Parameter Estimates:Store-Switching (Controls: (1) City Change, (2) Partner Store)

	(1)	(2)
purchase_val	-0.082**	-0.100**
	(0.003)	(0.003)
delay_std	0.020**	0.021**
	(0.002)	(0.002)
experience	-0.212**	-0.207**
	(0.002)	(0.002)
missing_std	0.140**	0.133**
	(0.002)	(0.002)
substitutions_std	0.140**	0.115**
	(0.002)	(0.002)
city_change	0.167**	
	(0.016)	
partner_store		-0.401**
		(0.009)
constant	-0.082**	0.294**
	(0.002)	(0.008)
Observations	435,737	434,240

** $p < 0.05$

Table C.4: MH Sampling Parameter Estimates:Store-Switching (User RE; Red. Sample)

	(1)
purchase_val	-0.203**
	(0.012)
delay_std	0.138**
	(0.067)
experience	-0.223**
	(0.036)
missing_std	0.277**
	(0.035)
substitutions_std	0.107**
	(0.059)
constant	-0.057**
	(0.022)
var(_cons[user_id])	0.076***
	(0.041)
Observations	73,718

** $p < 0.05$

Table C.5: ME Probit model for Store-switching (excluding platform-exit)

	(1)	(2)	(3)
purchase_val	-0.090***	-0.091***	-0.091***
	(0.004)	(0.004)	(0.004)
missing_std	0.130***	0.127***	0.130***
	(0.003)	(0.003)	(0.003)
delay_std	0.020***	0.020***	0.020***
	(0.002)	(0.002)	(0.003)
experience	-0.150***	-0.161***	-0.162***
	(0.003)	(0.003)	(0.003)
substitutions_std	0.131***	0.127***	0.128***
	(0.003)	(0.003)	(0.003)
missing_std*substitutions_std		-0.003	
		(0.003)	
missing_std*experience		-0.017***	-0.017***
		(0.004)	(0.004)
substitutions_std*experience		-0.026***	
		(0.003)	
missing_std*substitutions_std*experience		0.001	
		(0.004)	
missing_std*delay_std			-0.012***
			(0.002)
delay_std*experience			-0.013***
			(0.004)
delay_std*substitutions_std			-0.003
			(0.003)
missing_std*delay_std*experience			-0.002
			(0.003)
substitutions_std*delay_std*experience			0.003
			(0.004)
missing_std*delay_std*substitutions_std			-0.003
			(0.002)
missing_std*delay_std*substitutions_std*experience			-0.006*
			(0.004)
constant	-0.137***	-0.142***	-0.142***
	(0.003)	(0.003)	(0.004)
var(_cons[user_id])	0.391***	0.390***	0.390***
	(0.005)	(0.005)	(0.005)
Observations	394,657	394,657	394,657

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table C.6: MH Sampling Parameter Estimates: Store-switch (excluding platform-exit)

	(1)	(2)
purchase_val	-0.081**	-0.081**
	(0.003)	(0.003)
delay_std	0.020**	0.019**
	(0.002)	(0.002)
experience	-0.173**	-0.180**
	(0.002)	(0.002)
missing_std	0.134**	0.132**
	(0.002)	(0.002)
substitutions_std	0.130**	0.127**
	(0.002)	(0.002)
missing_std*experience		-0.009**
		(0.003)
substitutions_std*experience		-0.025**
		(0.002)
missing_std*substitutions_std		-0.009**
		(0.002)
missing_std*substitutions_std*experience		0.002
		(0.002)
constant	-0.216**	-0.219**
	(0.002)	(0.002)
Observations	394,657	394,657

** $p < 0.05$

Table C.7: MH Sampling Parameter Estimates: Platform-Exit (1-month cut)

	(1)	(2)
purchase_val	-0.135**	-0.136**
	(0.003)	(0.003)
delay_std	0.008**	0.007**
	(0.002)	(0.002)
experience	-0.519**	-0.510**
	(0.004)	(0.004)
missing_std	0.087**	0.095**
	(0.002)	(0.002)
substitutions_std	0.075**	0.082**
	(0.002)	(0.003)
missing_std*experience		0.022**
		(0.005)
substitutions_std*experience		0.020**
		(0.004)
missing_std*substitutions_std		0.002
		(0.002)
missing_std*substitutions_std*experience		0.015**
		(0.005)
constant	-0.659**	-0.655**
	(0.002)	(0.002)
Observations	448,537	448,537

** $p < 0.05$

Table C.8: MH Sampling Parameter Estimates: Platform-Exit (2-month cut; Recent store-switch control)

	(1)	(2)
purchase_val	-0.081**	-0.081**
	(0.003)	(0.003)
delay_std	0.027**	0.027**
	(0.002)	(0.002)
experience	-0.719**	-0.719**
	(0.006)	(0.006)
missing_std	0.104**	0.104**
	(0.003)	(0.003)
substitutions_std	0.093**	0.093**
	(0.003)	(0.003)
store_switch	0.288**	0.288**
	(0.005)	(0.005)
missing_std*experience		0.018**
		(0.005)
substitutions_std*experience		0.022**
		(0.005)
missing_std*substitutions_std		-0.000
		(0.003)
missing_std*substitutions_std*experience		0.014**
		(0.006)
constant	-1.040**	-0.977**
	(0.004)	(0.003)
Observations	383,089	383,089

** $p < 0.05$

Table C.9: Multinomial Logit model: With user demographics

	Roma	Milano	Torino
purchase_val	0.362***	1.619***	6.757***
	(0.008)	(0.030)	(0.162)
stockout_std	0.680***	1.723***	0.883***
	(0.015)	(0.039)	(0.015)
substitutions_std	0.377***	1.723***	1.767***
	(0.009)	(0.043)	(0.036)
stockout_std*substitutions_std	1.300***	0.940***	0.943***
	(0.014)	(0.009)	(0.008)
experience	0.786***	0.486***	0.799***
	(0.015)	(0.011)	(0.012)
stockout_std*experience	0.840***	1.208***	0.988
	(0.026)	(0.032)	(0.020)
substitutions_std*experience	0.590***	1.336***	1.309***
	(0.021)	(0.041)	(0.031)
stockout_std*substitutions_std*experience	1.225***	0.960***	0.951***
	(0.023)	(0.012)	(0.011)
delay_std	1.002	0.825***	1.091***
	(0.011)	(0.011)	(0.013)
gender	1.236***	0.742***	0.859***
	(0.031)	(0.024)	(0.023)
age	0.999**	1.002***	1.000
	(0.000)	(0.000)	(0.000)
15			
purchase_val	2.257***	0.463***	
	(0.047)	(0.021)	
stockout_std	1.935***	0.967	
	(0.056)	(0.041)	
substitutions_std	1.657***	1.057	
	(0.055)	(0.048)	
stockout_std*substitutions_std	0.956***	1.015	
	(0.010)	(0.022)	
experience	0.121***	0.704***	
	(0.008)	(0.024)	
stockout_std*experience	1.646***	0.917	
	(0.081)	(0.055)	
substitutions_std*experience	1.415***	0.932	
	(0.080)	(0.059)	
stockout_std*substitutions_std*experience	0.953***	1.002	
	(0.017)	(0.028)	

(Cont'd on following page)

Table C.9, continued

	Roma	Milano	Torino
delay_std	0.813***	0.745***	
	(0.016)	(0.018)	
gender	0.787***	1.271***	
	(0.034)	(0.056)	
age	1.001***	1.000	
	(0.000)	(0.000)	
30			
purchase_val	1.983***		
	(0.052)		
stockout_std	1.472***		
	(0.049)		
substitutions_std	1.097**		
	(0.043)		
stockout_std*substitutions_std	0.994		
	(0.013)		
experience	0.441***		
	(0.020)		
stockout_std*experience	1.236***		
	(0.065)		
substitutions_std*experience	1.240***		
	(0.073)		
stockout_std*substitutions_std*experience	0.955**		
	(0.020)		
delay_std	0.768***		
	(0.022)		
gender	0.815***		
	(0.045)		
age	1.000		
	(0.000)		
50			
purchase_val	0.342***	0.115***	
	(0.008)	(0.003)	
stockout_std	1.232***	1.626***	
	(0.023)	(0.029)	
substitutions_std	0.449***	0.645***	
	(0.010)	(0.013)	
stockout_std*substitutions_std	1.060***	1.010	
	(0.014)	(0.010)	
experience	0.873***	1.280***	
	(0.016)	(0.015)	

(Cont'd on following page)

Table C.9, continued

	Roma	Milano	Torino
stockout_std*experience	1.085***	0.979	
	(0.027)	(0.018)	
substitutions_std*experience	0.660***	0.929***	
	(0.020)	(0.020)	
stockout_std*substitutions_std*experience	1.105***	1.001	
	(0.023)	(0.012)	
delay_std	0.954***	0.802***	
	(0.010)	(0.008)	
gender	1.146***	1.007	
	(0.029)	(0.022)	
age	1.000*	1.002***	
	(0.000)	(0.000)	
51			
purchase_val	1.920***	1.184***	6.690***
	(0.027)	(0.022)	(0.187)
stockout_std	0.771***	1.351***	1.001
	(0.013)	(0.030)	(0.024)
substitutions_std	0.932***	1.035	1.535***
	(0.017)	(0.026)	(0.043)
stockout_std*substitutions_std	1.078***	1.011	0.964***
	(0.008)	(0.010)	(0.010)
experience	0.827***	0.635***	0.611***
	(0.009)	(0.011)	(0.015)
stockout_std*experience	1.004	1.051*	1.088***
	(0.019)	(0.028)	(0.032)
substitutions_std*experience	0.896***	1.018	1.296***
	(0.018)	(0.032)	(0.047)
stockout_std*substitutions_std*experience	1.054***	1.005	0.944***
	(0.011)	(0.013)	(0.014)
delay_std	0.957***	0.799***	1.087***
	(0.008)	(0.010)	(0.018)
gender	0.869***	0.964	0.923**
	(0.019)	(0.027)	(0.037)
age	1.001***	1.000*	1.001***
	(0.000)	(0.000)	(0.000)
60			
purchase_val	1.168***		
	(0.038)		
stockout_std	1.021		
	(0.035)		

(Cont'd on following page)

Table C.9, continued

	Roma	Milano	Torino
substitutions_std	0.928*		
	(0.036)		
stockout_std*substitutions_std	0.977		
	(0.020)		
experience	0.596***		
	(0.020)		
stockout_std*experience	0.988		
	(0.055)		
substitutions_std*experience	1.160**		
	(0.072)		
stockout_std*substitutions_std*experience	0.911***		
	(0.028)		
delay_std	1.027		
	(0.018)		
gender	1.179***		
	(0.053)		
age	1.000		
	(0.000)		
69			
purchase_val	0.798***		
	(0.029)		
stockout_std	1.973***		
	(0.053)		
substitutions_std	0.953		
	(0.033)		
stockout_std*substitutions_std	0.938***		
	(0.014)		
experience	0.630***		
	(0.024)		
stockout_std*experience	1.427***		
	(0.052)		
substitutions_std*experience	0.961		
	(0.050)		
stockout_std*substitutions_std*experience	0.984		
	(0.023)		
delay_std	0.901***		
	(0.017)		
gender	1.112**		
	(0.052)		
age	1.000		

(Cont'd on following page)

Table C.9, continued

	Roma	Milano	Torino
	(0.000)		
70			
purchase_val	0.974		
	(0.030)		
stockout_std	0.907***		
	(0.031)		
substitutions_std	0.774***		
	(0.030)		
stockout_std*substitutions_std	1.010		
	(0.021)		
experience	1.021		
	(0.018)		
stockout_std*experience	1.068**		
	(0.034)		
substitutions_std*experience	0.814***		
	(0.030)		
stockout_std*substitutions_std*experience	0.986		
	(0.027)		
delay_std	0.902***		
	(0.019)		
gender	0.993		
	(0.044)		
age	1.000		
	(0.000)		
204			
purchase_val	0.930**		
	(0.028)		
stockout_std	1.481***		
	(0.039)		
substitutions_std	0.953		
	(0.030)		
stockout_std*substitutions_std	0.977*		
	(0.013)		
experience	0.617***		
	(0.019)		
stockout_std*experience	1.168***		
	(0.047)		
substitutions_std*experience	1.117**		
	(0.054)		
stockout_std*substitutions_std*experience	0.962*		

(Cont'd on following page)

Table C.9, continued

	Roma	Milano	Torino
	(0.021)		
delay_std	0.811***		
	(0.016)		
gender	0.998		
	(0.040)		
age	1.001		
	(0.000)		
7			
purchase_val		0.961	5.882***
		(0.024)	(0.161)
stockout_std		1.298***	0.835***
		(0.035)	(0.018)
substitutions_std		1.090***	1.751***
		(0.033)	(0.044)
stockout_std*substitutions_std		0.959***	0.928***
		(0.013)	(0.011)
experience		0.662***	0.648***
		(0.015)	(0.014)
stockout_std*experience		1.160***	1.033
		(0.038)	(0.029)
substitutions_std*experience		0.985	1.398***
		(0.038)	(0.046)
stockout_std*substitutions_std*experience		0.952***	0.921***
		(0.017)	(0.014)
delay_std		0.703***	0.992
		(0.013)	(0.016)
gender		0.982	0.945*
		(0.033)	(0.031)
age		1.000	1.001**
		(0.000)	(0.000)
12			
purchase_val		0.674***	
		(0.017)	
stockout_std		1.624***	
		(0.036)	
substitutions_std		1.327***	
		(0.034)	
stockout_std*substitutions_std		0.887***	
		(0.010)	
experience		1.122***	

(Cont'd on following page)

Table C.9, continued

	Roma	Milano	Torino
		(0.016)	
stockout_std*experience		1.005	
		(0.022)	
substitutions_std*experience		1.285***	
		(0.033)	
stockout_std*substitutions_std*experience		0.926***	
		(0.012)	
delay_std		0.927***	
		(0.011)	
gender		0.864***	
		(0.026)	
age		1.000*	
		(0.000)	
40			
purchase_val		0.617***	5.491***
		(0.032)	(0.131)
stockout_std		0.774***	0.765***
		(0.046)	(0.013)
substitutions_std		1.426***	1.679***
		(0.086)	(0.033)
stockout_std*substitutions_std		0.940**	0.956***
		(0.028)	(0.008)
experience		1.163***	0.824***
		(0.033)	(0.012)
stockout_std*experience		0.814***	0.935***
		(0.049)	(0.019)
substitutions_std*experience		1.194***	1.291***
		(0.072)	(0.030)
stockout_std*substitutions_std*experience		0.967	0.955***
		(0.033)	(0.011)
delay_std		0.774***	1.134***
		(0.024)	(0.012)
gender		0.808***	0.900***
		(0.050)	(0.023)
age		1.000	1.000
		(0.000)	(0.000)
49			
purchase_val		0.960**	
		(0.020)	
stockout_std		1.088***	

(Cont'd on following page)

Table C.9, continued

	Roma	Milano	Torino
		(0.027)	
substitutions_std		0.913***	
		(0.024)	
stockout_std*substitutions_std		1.058***	
		(0.011)	
experience		0.856***	
		(0.014)	
stockout_std*experience		1.038	
		(0.026)	
substitutions_std*experience		0.939**	
		(0.027)	
stockout_std*substitutions_std*experience		1.043***	
		(0.013)	
delay_std		0.883***	
		(0.011)	
gender		1.086***	
		(0.031)	
age		1.002***	
		(0.000)	
54			
purchase_val		0.038***	0.299***
		(0.002)	(0.025)
stockout_std		2.142***	1.292***
		(0.068)	(0.055)
substitutions_std		0.828***	1.141**
		(0.033)	(0.061)
stockout_std*substitutions_std		1.010	1.006
		(0.014)	(0.016)
experience		1.157***	0.766***
		(0.037)	(0.042)
stockout_std*experience		0.998	1.023
		(0.042)	(0.065)
substitutions_std*experience		1.072	1.098
		(0.056)	(0.087)
stockout_std*substitutions_std*experience		1.038**	1.043
		(0.018)	(0.027)
delay_std		0.807***	1.056**
		(0.017)	(0.027)
gender		1.212***	0.897*
		(0.056)	(0.058)

(Cont'd on following page)

Table C.9, continued

	Roma	Milano	Torino
age		1.002***	1.001
		(0.000)	(0.001)
5			
purchase_val			3.432***
			(0.077)
stockout_std			0.524***
			(0.008)
substitutions_std			1.834***
			(0.032)
stockout_std*substitutions_std			0.944***
			(0.008)
experience			0.958***
			(0.012)
stockout_std*experience			0.829***
			(0.016)
substitutions_std*experience			1.317***
			(0.028)
stockout_std*substitutions_std*experience			0.930***
			(0.011)
delay_std			1.008
			(0.010)
gender			0.806***
			(0.017)
age			1.000**
			(0.000)
Observations	104105	103556	88168
Pseudo R2	0.099	0.091	0.057

Figures

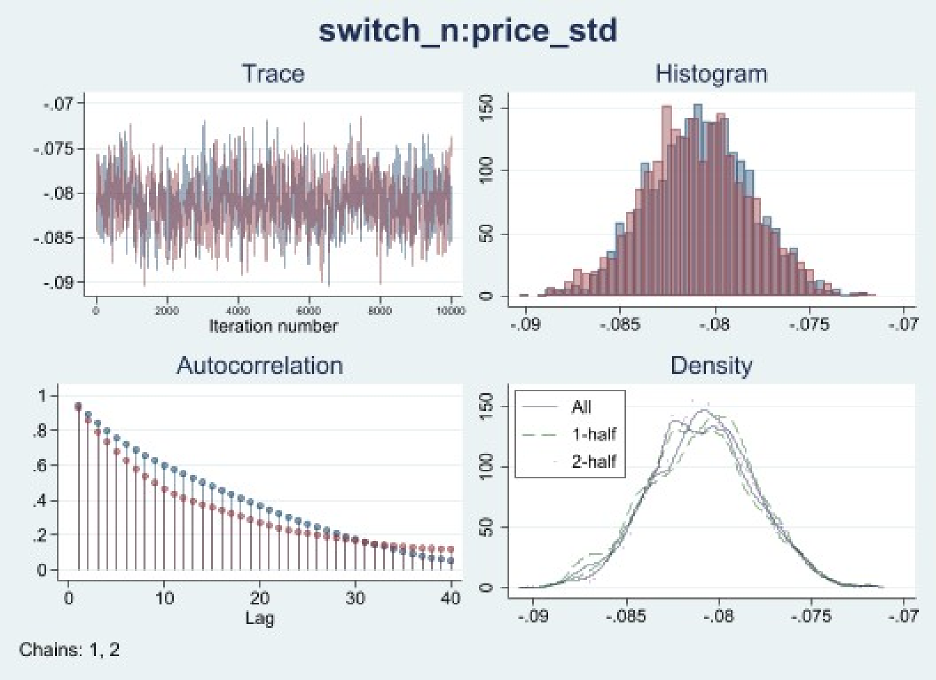


Figure C.1: Plot for Switch:Purchase Value (Excluding platform-exit)

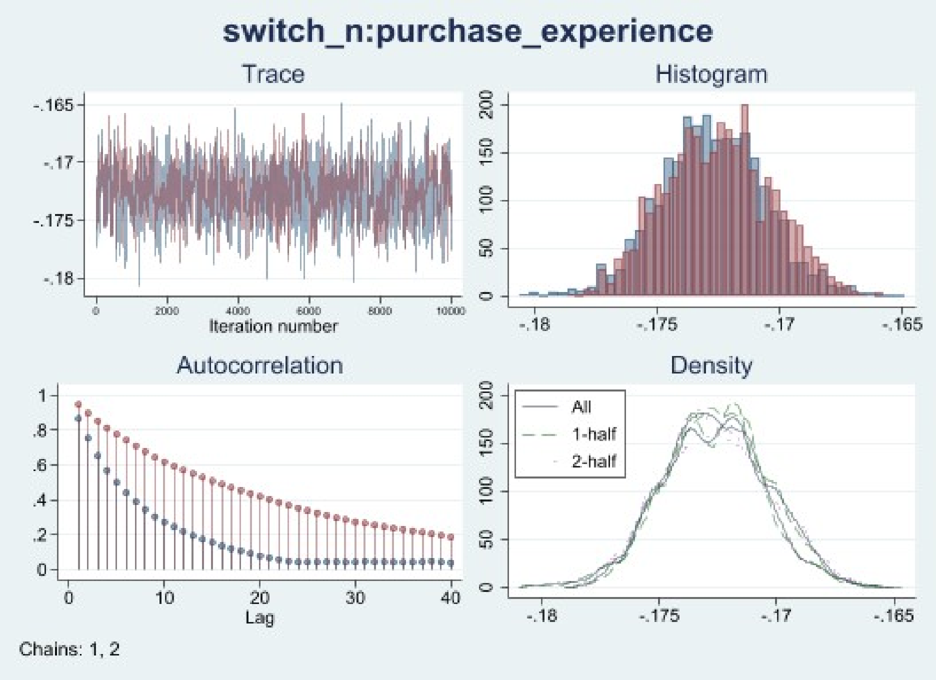


Figure C.2: Plot for Switch:Experience (Excluding platform-exit)

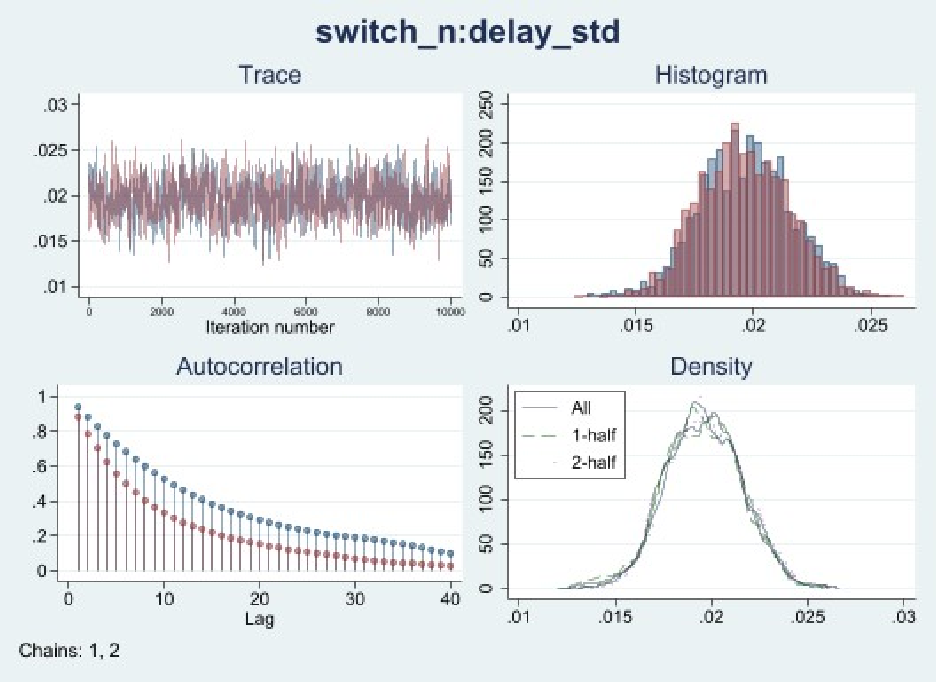


Figure C.3: Plot for Switch:Delay (Excluding platform-exit)

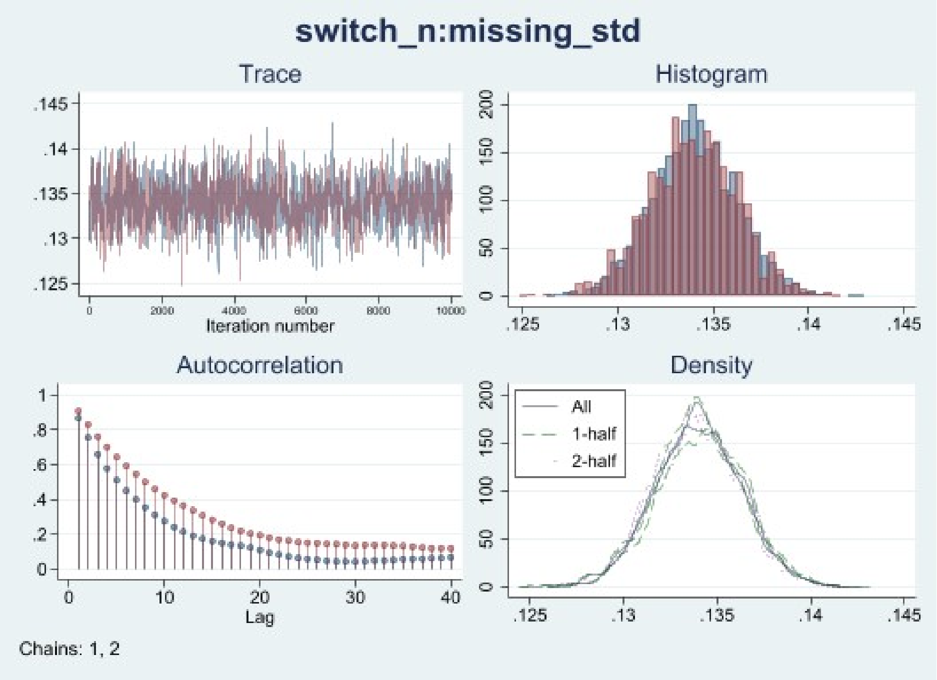


Figure C.4: Plot for Switch:Missing (Excluding platform-exit)

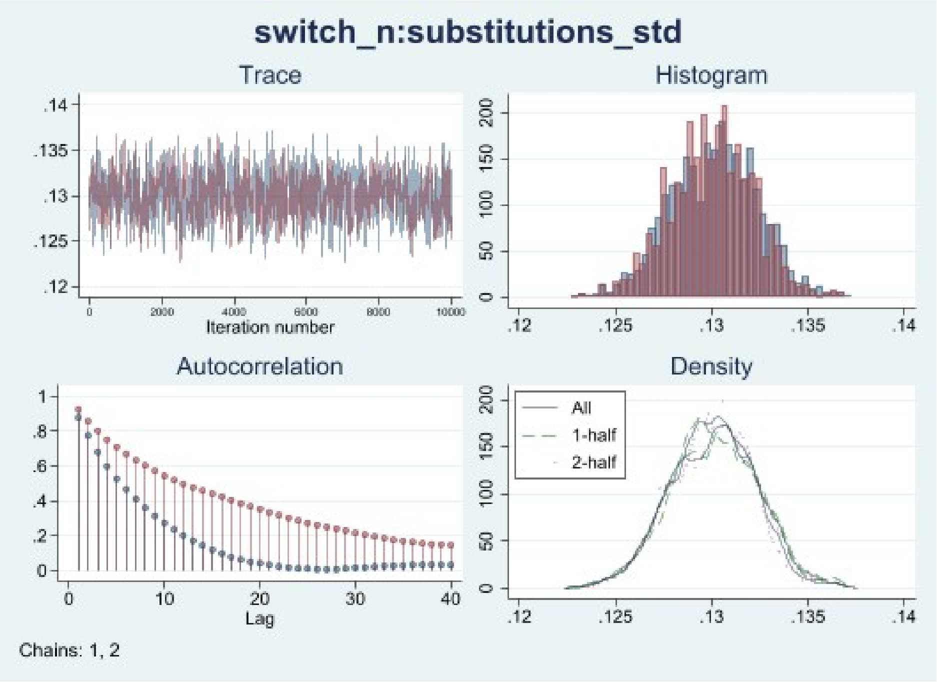


Figure C.5: Plot for Switch:Substitutions (Excluding platform-exit)