

Impact of Task-Level Worker Specialization, Workload, and Product Personalization on Consumer Returns

Hailong Cui^a • Sampath Rajagopalan^b • Amy R. Ward^c

^{ab}Marshall School of Business, University of Southern California ^cBooth School of Business, University of Chicago

hailongc@marshall.usc.edu • raj@marshall.usc.edu • amy.ward@chicagobooth.edu

August 10, 2019

Abstract

Problem Definition: Are consumer product returns largely a function of retailers' return policies or can manufacturers influence them through production characteristics and product attributes? How do operational levers under a manufacturer's control impact return rates?

Academic/Practical Relevance: Consumer returns raise major questions for manufacturers and retailers because losses due to returns are substantial. However, levers commonly used to discourage returns, such as strict return policies, can dampen sales, which is a significant drawback. The literature on consumer returns has focused on retailer and purchase attributes but has ignored whether manufacturers can take a proactive role to influence returns. Manufacturers and retailers have an interest in understanding how operational levers influence return rates.

Methodology: This work attempts to answer questions about the influence of manufacturing on return rates using detailed operational data from a leading U.S. manufacturer of automotive accessories. We study the impact of key operational levers on return rates using a logistic regression model, while controlling for retailer return policies and numerous product, purchase, and consumer attributes.

Results: We find that increased task-level worker specialization has a significant and U-shaped impact on return rates. Increased workload levels in production leads to a significant increase in return rates whereas product personalization leads to lower return rates.

Managerial Implications: We provide an empirical estimate of the effect of operational variables on return rates that are useful in evaluating managerial trade-offs. Our findings suggest that allowing customers to personalize products can significantly reduce return rates.

Keywords: Production Process, Personalization, Returns, Econometrics, Marketing–Operations Interface.

1. Introduction

Consumer product returns are a major issue for retailers and manufacturers. Returned merchandise in the U.S. was estimated at \$369 billion out of sales of \$3.7 trillion in 2018 (Appriss Retail 2018), an increase in both return volume and return rate from \$261 billion out of \$3.3 trillion in 2014 (The Economist 2014). E-commerce experiences a higher return rate than brick-and-mortar stores; 25% of all goods bought online are estimated to be returned (Adams 2017) even though the return rate can vary substantially across different retail categories (Appriss Retail 2018). Furthermore, more than half of all returns may not be resold at full

price, which results in substantial financial losses for retailers and manufacturers (Cheng 2015). Returns may be especially costly for manufacturers if retailers simply pass along the returned goods to the manufacturers. Returns are also important because they are an important driver of customer sentiment, which impacts a firm’s reputation, especially in e-commerce (Barnhar 2019). While the costs of a high return rate have drawn attention to the management of returned goods in both the media and academic research (e.g., Howland 2018, Guide et al. 2006), the manufacturer’s role in reducing returns has received little attention beyond anecdotal accounts.

On the retailer’s side, some firms are taking measures to reduce returns. For example, L.L. Bean is dropping the liberal return policy that was a key part of its value proposition for decades and Amazon has begun to ban certain shoppers from their site for excessive returns (Dennis 2018, Safdar and Stevens 2018). Firms charge restocking fees and institute other policies that may help reduce returns. However, many firms are reluctant to institute strict return policies because while they may deter returns, they also may dampen sales (Anderson et al. 2009). The marketing literature has already identified many factors that influence returns, including product attributes (e.g., fit issues), purchase attributes (e.g., price, promotion), and retailer attributes (e.g., return policies, large store network) (see Petersen and Kumar 2009 and references therein). The literature, however, has not sufficiently studied factors that influence returns on the manufacturer’s side.

On the manufacturer’s side, returns directly affect a firm’s bottom line because they are often not resalable or must be sold at a lower price or a loss. Also, unlike the case with retailer and purchase attributes, manufacturers have greater control over several aspects of the production characteristics. Thus, improving the production process may lead to better quality, and in turn can help reduce returns, but does not apparently face the trade-off between sales and returns that a retailer faces when it implements a stricter return policy.

In order to fill that gap in the literature, the primary objective of this paper is to empirically investigate the impact of production process-related factors on returns. Unlike prior studies focusing on retailers’ return policies and their impact on returns, we study how key operational levers available to manufacturers affect product returns, while controlling for different return policies and numerous product, purchase, and consumer attributes. In particular, we investigate the impact of two production variables on returns, viz. worker specialization (on identical or similar tasks) and production workload, both of which have been identified as important factors that influence productivity and quality in the Operations Management (OM) literature. The impact of these production variables is of interest to manufacturers of consumer products with labor-intensive production processes such as apparel and automobile accessories. We also study the impact of product personalization (e.g., embroidering the logo of a buyer’s alma mater in a car seat cover) on product returns because the psychology literature points out that personalization may increase a consumer’s attachment to a product (see Mugge et al. 2009 for an experimental study) and because a manufacturer can easily offer such personalization. In fact, personalization is expected to be the prime driver of marketing success within five years (McKinsey 2019). To the best of our knowledge, no empirical study has explored

the impact of these operational levers on returns. In addition, we include several control variables that have not been considered in prior studies on product returns.

To explore the above issues, we collected data from a leading U.S. manufacturer of automotive accessories, which primarily sells seat covers, vehicle covers, and dash covers through various online retailers. The company sells products that are tailored to individual vehicle makes and models, and offers them in various fabric types and styles. Given the vast variety of offerings, the firm must manufacture products on a make-to-order basis and aims to ship them to consumers in 1 to 2 weeks. Production of the products requires several sewers, so the process design and level of worker specialization are critical issues for the firm. Moreover, due to the make-to-order production, management of workload is important from a production and lead time perspective. Finally, the firm has a very good Enterprise Resource Planning (ERP) system that tracks various aspects of every single item sold from order to delivery as well as possible return, including which sewers have worked on the product. For these reasons, this firm presents a good setting for our study.

We study the drivers of return rates using a logistic regression model wherein the dependent variable is the probability of a return. We find that worker specialization is a significant factor that impacts the probability of a return, even after controlling for numerous product, retailer, and purchase attributes. Interestingly, the impact of worker specialization is U-shaped; that is, the probability of a return first decreases over a range of specialization values, and then increases after specialization reaches a certain threshold. We also find that the probability of a return increases with workload, plausibly because higher workload may result in workers rushing and producing more defects. Finally, we find that product personalization in the form of logos reduces the consumer's propensity to return a product. The effect of personalization on return rates is large and significant, while the effects of specialization and workload are small despite being significant. We demonstrate that the findings are robust by exploring alternative measures for worker specialization and workload, changing some control variables, dropping a subset of the data, and analyzing potential endogeneity issues.

Managers should find our result on personalization interesting because it suggests that manufacturers can reduce product return rates substantially by allowing the customers to personalize the product. In addition, our results suggest that achieving the right level of worker specialization and decreasing workload levels may help reduce return rates, but the effects are relatively small and managers have to take into account the cost implications of their actions. These drivers may be more attractive than strict return policies because they will not dampen sales and may, in fact, increase sales in the long run because of the reputation gained from better quality and personalized product offerings.

The remainder of the paper is organized as follows. We first discuss the related literature and highlight our contributions relative to this literature. Then, in §2 we discuss the empirical setting and the data used in our study. In §3, we develop the hypotheses, and discuss the relevant measures and the control variables used. In §4 we discuss the estimation strategy, results, and robustness issues for estimating the return rate. We conclude with a discussion in §5 and provide additional tables in the online appendix.

Literature Review

This study is related to three main streams of literature: (1) research in marketing and information systems on consumer returns, (2) research in OM on returns, (3) empirical studies in OM on the impact of specialization and workload.

The literature in marketing on product returns focuses on the influence of consumer, retailer, and product attributes on returns but does not consider the impact of a manufacturer’s production process. Hess and Mayhew (1997) is among the earliest works to empirically study factors that impact returns. They develop a hazard model for explaining and predicting the return rate and time to return using price and importance of fit as covariates and test them on data from a catalog company. They find price to be significant in explaining return rates. Anderson et al. (2006) utilize an economic model of consumer purchase and returns behavior and find support for the perceived value hypothesis, which predicts that consumer return rates increase with the price paid. Anderson et al. (2009) develop a structural model of a consumer’s decision to purchase and return a product, which enables a retailer to both measure the value to consumers of the return option and assess the costs and benefits of different return policies. In an experimental study, Wood (2001) finds that lenient policies do not necessarily increase returns due to the endowment effect. More recently, Janakiraman et al. (2016) show that overall, return policy leniency increases purchases more than returns. Petersen and Kumar (2009) determine the factors in the firm-consumer exchange process that help explain product return behavior and the consequences of product returns on future consumer and firm behavior. Further, Petersen and Kumar (2015) show that a firm is able to increase both its short- and long-term profits when accounting for the consumer’s perceived risk related to product returns in addition to managing product return costs. Hong and Pavlou (2014) explore the impact of product fit uncertainty and quality uncertainty on product returns and customer satisfaction in online markets; data on returns, product fit and quality uncertainty for this empirical study were obtained from the consumers of major online sites using surveys.

In contrast to the aforementioned research, our paper explicitly studies the impact of a manufacturer’s production process to explore how it can use its operational levers to influence the return rate, while controlling for product, consumer, and retailer characteristics. For instance, unlike the quality uncertainty studied in Hong and Pavlou (2014), our focus is on the impact of operational levers on returns, where those operational levers can impact quality, and we use archival data rather than surveys.

The OM literature on product returns is largely analytical in nature with the goal of using models to understand how different dimensions of retailer policies may impact returns—for example, restocking fees (Shulman et al. 2011) and return policies (Su 2008, Altug and Aydinliyim 2016). Lu and Chen (2012) study analytically the impact of product quality and other key variables on a firm’s return policy, including restocking fees. They show interestingly that higher product quality need not imply a more generous return policy. Some recent works also conduct empirical studies on returns from the perspective of retailers and consumers. For example, Shang et al. (2017) study both the return policy drivers from the retailer’s perspective and the return policy value from the consumer’s perspective, and show that the value of a full refund

policy to consumers may not be as large as one might expect. Rao et al. (2004) consider the impact on product returns of disclosure of inventory availability prior to purchasing products and the post-sale delivery reliability of purchased goods in online retail sales. Mollenkopf et al. (2007) study the impact of the returns management system (web interface, service quality, reverse logistics transactional flow, etc.) upon customer loyalty intentions using surveys of online customers. Griffis et al. (2012) study the relationship between a customer's experience of product returns, and subsequent shopping behavior using actual purchase and returns history at an online retailer. There is a related stream of literature in OM that studies the value recovery from returned products by remanufacturing and issues related to return logistics (see Guide et al. 2006, Pınç et al. 2016 and references therein).

In contrast to the aforementioned research, our study is empirical (not analytical) in nature, focuses on the manufacturer's (not retailer's or consumer's) perspective, and is not concerned with remanufacturing or reverse logistics.

Specialization and its impact on productivity have been of interest to managers and researchers for a long time, ever since Adam Smith and Frederick Taylor (Taylor 1911). More recent works have studied other potential benefits and costs of specialization. For example, in the OM literature, Staats and Gino (2012) investigate the effect of specialization in conjunction with variety on short- and long-term worker productivity in the completion of simple, repetitive tasks in a banking operation. To our knowledge, there have been very few studies that have explored the impact of specialization on quality, which could then influence returns. A somewhat related study is KC (2014) which finds that multitasking by physicians in a hospital emergency department results in poorer quality as measured by higher patient readmission rates. Our focus is on the impact of specialization and process design, and not on the impact of multitasking by an individual worker. In our data, we can see, for a product, how many similar tasks each contributing worker has performed in the past, which can be different for different types of products and materials. This allows us to explicitly test for the significance of task-level worker specialization on product returns.

There is some empirical work on the impact of product *focus* (or limited variety) on quality. In early work on the auto assembly operations, Fisher and Ittner (1999) find that greater product variety has a significant adverse impact on minor repair and major rework. More recently, Shah et al. (2017) study the impact of product variety, plant variety, and capacity utilization on product recalls in the auto industry. Neither study explores the impact of task-level worker specialization on product returns. Even if variety is unchanged, specialization levels can vary across workers and over time depending on how tasks are assigned. KC and Terwiesch (2011) examine three distinct levels of the organization (firm, operating unit, and process flow) and find that focus at each of these levels is associated with higher levels of quality.

Workload, the second production process variable we study, has received a lot of attention in the OM literature especially in health-care settings. Some examples of recent empirical works include studies on the impact of physician workload on hospital reimbursement (Powell et al. 2012), impact of workload on the service time and patient outcomes in hospital operations (KC and Terwiesch 2009), and impact of workload

on sales and meal duration in a restaurant operation (Tan and Netessine 2014). Unlike these studies, we investigate the impact of workload, in conjunction with task-level worker specialization, and product personalization, on consumer returns. The relatively small effect size of workload in our study is similar to the finding in KC and Terwiesch (2009) that increased workload can result in early patient discharges which results in a small increase in the post-discharge mortality rate.

Another important lever that a manufacturer can use to impact return rates is product personalization, which has been shown to increase a consumer’s interest in and attachment to a product according to the consumer psychology literature. Mugge et al. (2009), in an experimental study, find that personalizing a product’s appearance increases the emotional bond with the product from a direct effect (as a result of the extended period of time spent with the product) and an indirect effect (via the personalized product’s self-expressive value). Howard and Kerin (2004) conducted field experiments and find that direct mail persuasion with a hand-written note with the recipient’s name increased free sample requests when strong brand attributes were used. Our study appears to be the first one to investigate empirically the impact of product personalization on return rates.

2. Empirical Setting

We first discuss the operational context and process for the study in §2.1 and then describe the data in §2.2.

2.1 Operational Context and Process

Our data come from a manufacturer of make-to-order automotive accessories (referenced as Company *A* hereafter). We focus on three distinctly different product types manufactured in the U.S. and shipped to U.S. consumers—vehicle cover, seat cover, and dash cover. After receipt of an order, Company *A* manufactures and ships its products to consumers in around one to two weeks. Due to the nature of make-to-order products, the retailers serve as online sales channels and do not carry any inventory of finished goods.

We visited their factory location in California and observed actual operations on multiple occasions between 2015 and 2018. We also met with the management to understand different product manufacturing processes, and how data is recorded in the ERP system.

Consumer Ordering Process: A consumer visits an online retailer, finds an automotive accessory that she likes, and submits an order. She needs to know basic information about the vehicle for which she orders the accessory, including make, model, year, and trim level. Once she has input information for her vehicle, she can choose the product she wants and specify fabric type, color, design, and whether to have a personalized logo. Sometimes, a consumer may request multiple products in the same order. After an order is submitted to the retailer, the information is transferred to Company *A*, which manufactures and ships the product to the consumer directly.

Manufacturing Process: Order labels are printed in response to customer order arrivals, and contain the detailed product(s) information. The production manager either immediately releases the order to the factory floor, or requests a delay, depending on current workload on the floor.

The manufacturing process can be broadly categorized into three stages: pre-sewing, sewing, and post-sewing. The equipment and workers at the three stages are in distinct, demarcated areas of an open factory floor. We provide an overview of each stage next, but discuss process details related to the variables of interest in our study in greater depth in §3. Afterwards, we discuss in detail work-in-process (WIP), workers, and product defects which could impact returns.

The *pre-sewing stage* starts with printing the order label, which contains detailed instructions for each of the subsequent operations. An employee in charge of cutting fabric picks up the required fabric of certain type and color from raw material inventory based on the order label and then brings the fabric to one of the available computerized cutting machines. She then scans the order label to add the cutting instructions to the machine, which automatically cuts the fabric. Afterwards, she inspects the cut fabric pieces, puts them in a plastic bag with the order label attached, and brings the bag to a common station with bins of cut fabric bundles. From this common station, an employee called a runner picks up the cut fabric bundles and *randomly* routes them to the individual inbound bins of different sewers.

Next, during the *sewing stage*, the orders are processed by sewers who perform sewing operations such as joining, binding, and embroidery. The sewers are grouped into three different production areas, each dedicated to one of the three product types mentioned earlier. The number of sewers assigned to each area is based upon the demand rate and processing rate for the products. Sewers typically work on the same product type but can be moved to a different product type depending on changes in demand rate or sewer availability. The number of tasks in sewing and the nature of the task vary with the product type. For example, dash covers and vehicle covers may require sewing a few small and large pieces of fabrics respectively, whereas seat covers often require sewing dozens of small pieces. Another source of task variation comes from the materials used which are categorized into three levels of complexity of sewing—low, medium, high in materials codes. We provide the names and descriptions for each task and the product types and material codes on which they can be performed in Table A1 in the online appendix.

Each sewer has an inbound bin and an outbound bin next to him that can hold a few units. The runner takes the bundles of cut fabric from the central station and places them in the inbound bin of the sewer and takes completed items from the outbound bin to another sewer if it requires additional sewing operations or to the packing stage. The runner assigns orders *randomly* to the sewers while ensuring that WIP at each sewer within a product pool is similar. Sewing is the most time-consuming and expensive part of the process (sewers are paid more than other workers) and is the bottleneck stage. The firm maintains a workforce level of sewers such that demand generally exceeds capacity so that sewers are always kept busy.

In the *post-sewing stage*, finished goods are packed and moved to the packing/shipping station, then shipped to the consumer. The employees in packing inspect every product for common and easily identified defects which include checking if the sewing is loose, material is damaged, and whether the order label description matches the product. However, this inspection cannot identify issues such as the fit of a car accessory to the vehicle.

WIP: There is WIP before the sewing stage, and there is WIP at each of the sewers, as discussed above. The total WIP of cut fabrics by product type before and during sewing can be obtained from the ERP. The post-sewing WIP that accumulates in the packing stage can also be derived from the ERP, but this is small. The WIP at each sewer is not recorded, but it is visible to the sewer.

Workers: The workers at the firm are paid on an hourly basis but a worker can also get a bonus based on their individual productivity which is tracked. So, there is some incentive to work faster. While this bonus may be an incentive to produce more output, a stronger incentive appears to be the impression workers make on management based on their productivity and quality. Anyone who performs poorly on a regular basis in terms of productivity and/or quality runs the risk of getting fired.

Management strongly believes that every sewer should be skilled at every task and product type. As a result, even though sewers are grouped into specific production areas by product type, any individual sewer may work on a varied set of products and perform a varied set of tasks. To ensure this flexibility, new workers, after a probationary period, typically produce dash or vehicle covers initially, and are then moved (after gaining adequate sewing experience), to seat covers, which are considered the most complex product type in terms of skill required.

Product Defects: Defects in the firm's products can be classified into two broad categories: material defects and production process defects. Examples of material defects include rips, fading over time, color variation, and broken zippers. Material related defects often get identified in-house but there are instances in which this is not true and they are identified by the consumer soon after purchase (color variation) or many months after purchase (faded dash cover). Examples of production-related defects include incorrect assembly, improper sewing, missing straps, and missing holes for a headrest in a seat cover. These defects may occur because a sewer is rushing, fatigued, or inexperienced. Some of these defects *cannot* be identified during inspection—for instance, (i) a car cover has to be fit on a car to identify that something is wrong or (ii) a missing hole in a seat cover for a headrest may not be identified because not all car seat covers need such holes. However, these defects are identified by a consumer soon after they attempt to use the product. The firm has not historically tracked the reason for returns and associated defects.

Returns Handling: A product may be returned within the return period for refund, or beyond the return period for repair or replacement under warranty. All the returns are handled by Company A, not the retailers. However, the return policy is set by each retailer and their return policy may have a strong influence on returns. In our data, we are unable to distinguish between returns due to production quality from returns due to non-quality issues (e.g., impulse purchase related returns) because quality is a latent variable and we do not have data on the reason for a return.

Return Policy of Retailers: For our analysis, we focus on the thirteen high-volume retailers that account for more than 90% of sales, and obtain their return policies on four common measures: return period, return authorization, restocking fees, and option to return to brick-and-mortar store. Details on the return policies of each of the thirteen retailers are provided in Table A2 in the online appendix. Each return

policy is based on a contract between each retailer and Company *A*, and none of the policies changed during the study period based on our discussion with the management at Company *A*.

2.2 Order Data Description

Company *A* utilizes the Microsoft ERP system to systematically track every operation performed on the order at the various stages that we discussed in the previous subsection, and Table 1 provides one example.

Table 1: A Sample Operational Data Set.

OrderNo	OrderDate	SerialNo	ScanDate	Retailer	ZIP Code [†]	ProductID	Material	Task	Logo	EmpID	ReturnDate
149038902	5-11-2015	L49325331	5-11-2015	xxx.com	90007	SCxxxx	High (Leather)	Printing Label	USC001	xx01	7-10-2015
149038902	5-11-2015	L49325331	5-14-2015	xxx.com	90007	SCxxxx	High (Leather)	Cutting Fabric	USC001	xx12	7-10-2015
149038902	5-11-2015	L49325331	5-15-2015	xxx.com	90007	SCxxxx	High (Leather)	Joining	USC001	xx53	7-10-2015
149038902	5-11-2015	L49325331	5-15-2015	xxx.com	90007	SCxxxx	High (Leather)	Binding	USC001	xx53	7-10-2015
149038902	5-11-2015	L49325331	5-18-2015	xxx.com	90007	SCxxxx	High (Leather)	Webbing Elastic	USC001	xx32	7-10-2015
149038902	5-11-2015	L49325331	5-18-2015	xxx.com	90007	SCxxxx	High (Leather)	Logo-Embroidery	USC001	xx26	7-10-2015
149038902	5-11-2015	L49325331	5-18-2015	xxx.com	90007	SCxxxx	High (Leather)	Packing	USC001	xx34	7-10-2015

[†] This column contains 5-digit ZIP Code (ZIP5) for the consumer’s shipping address.

The example shows two pre-sewing tasks (printing label, cutting fabric), four sewing tasks (joining, binding, webbing elastic, logo-embroidery), and one post-sewing task (packing) performed on a product with the serial number L49325331 for order number 149038902. We can identify the product type by the first two characters in the ProductID (SC stands for seat cover). The OrderDate column shows that this product was ordered on 5/11/2015 (Monday). In the ScanDate column, we can see that the order label was printed on the same day, but production was delayed until three days later 5/14/2015 (Thursday) when an employee read the label to find out which fabrics to pick up from inventory and cut the fabric in preparation for sewing according to the specification. This implies that the production manager did not immediately release the order into production due to high workload on the floor.

Sewers are often assigned to perform only a single task on a given product. However, we observe that on 5/15/2015 (Friday), an employee with ID xx53 performed two consecutive sewing operations—joining and binding. On 5/18/2015 (Monday), a different sewer added webbing elastic, and another sewer added a logo through embroidery and finished this product. The order label contains instructions for whether a sewer should perform one task or two or three consecutive tasks, taking into account various order characteristics. Depending on the completion time, this product was shipped to a consumer on the same day or the next day. The product was returned to the factory on 7/10/2015 as shown in the ReturnDate. As discussed, there is a large variability among tasks based on the product type, task names, and materials codes. We provide further details on the tasks in the Table A1 in the online appendix.

Our study is restricted to orders placed between 2013 and 2015 for a total of 179,906 products. Out of these, 15,088 were returned by the end of 2015 and 644 more were returned in 2016. These show that the manufacturer in our study experienced an average return rate of 8.74%. The return rate fluctuated by retailer between 4.65% and 12.92%, and varied by product type (9.52% in dash cover, 8.09% in seat cover, 6.40% in vehicle cover).

3. Hypotheses and Measures

In this section, we develop hypotheses related to the impact on return rates. The literature in marketing and operations has identified numerous factors that influence product return rates and we extend it to include manufacturing related factors that can impact returns. The marketing literature (e.g., Anderson et al. 2009, Petersen and Kumar 2015) adopts the notion that consumer returns are a result of a poor fit between the product and a consumer’s preferences and that fit is not fully observed by the consumer prior to purchase and receipt of the item. The information systems literature (e.g., Pavlou et al. 2007, Hong and Pavlou 2014) has identified product fit uncertainty and quality uncertainty as key drivers of product returns and customer satisfaction for experience goods. This notion is particularly valid in our context for several reasons. First, vehicle accessories are infrequently purchased products, similar to vehicles, so there is uncertainty about the fit and quality. Second, because the firm tries to customize the vehicle accessory to each specific vehicle, fit is especially important to a consumer in this context. Finally, because all the orders are placed online the consumer cannot physically assess the fit or the suitability of the product to their vehicle until after receipt of the item. The manufacturer may be able to influence quality via operational levers but not product fit uncertainty, and so the ability of the manufacturer to influence return rates is unclear from the aforementioned literature, which motivates further study.

In some of the OM literature, a similar approach theorizes that consumers return products because of valuation uncertainty, but that consumers can only determine a value, which depends on fit and quality, after receiving the product (e.g., Su 2008, Lu and Chen 2012). This approach adopts the framework that, having purchased a product, a return is more likely when the net value from return is greater than the net value from keeping the product. These net values are in turn impacted by numerous consumer and purchase characteristics (including the distribution channel and retailer through which the product is purchased) as well as specific factors that impact the return effort and cost. However, this literature does not explicitly consider the impact of product quality on the consumer’s propensity to return and production process characteristics that in turn influence quality. Our work explicitly considers the impact of production process characteristics on returns and also includes several new control variables that may impact return rates so as to reduce unobserved heterogeneity. Understanding the impact of these characteristics is important because the manufacturer has little direct control over consumer and purchase characteristics and return effort but does have control over the production processes that impact the quality.

The primary dependent variable in our study is the probability of a return—we use this term interchangeably with return rate—of a specific item purchased.

3.1 Factors Impacting Return Rates

Figure 1 provides a framework that identifies the numerous factors that influence return rates such as consumer, product, and purchase characteristics as well as return effort. These factors may impact the value of keeping the product or returning it or both. The quality of the product influences returns but is a latent

variable and cannot be observed directly. Moreover, the impact of quality must be separated from other factors that impact returns.

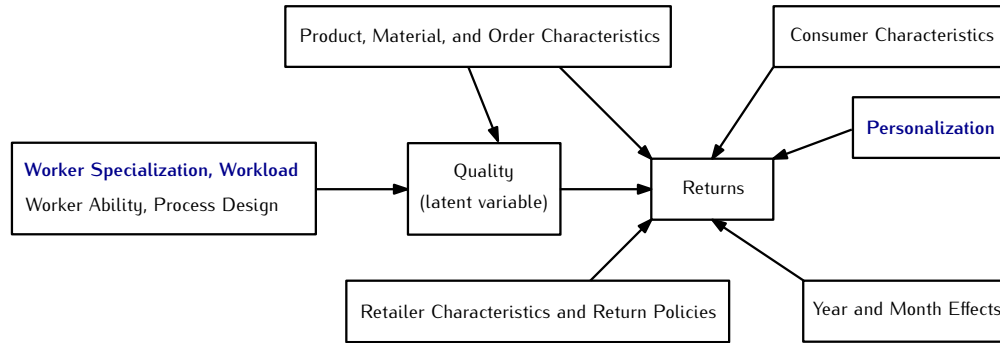


Figure 1: Factors Driving Consumer Product Returns.

We are interested in variables related to the production process that can influence quality which in turn impacts return rates. This is because identifying which (if any) of these variables are significant potentially allows the manufacturer to design the production process in a way that lowers return rates. We simultaneously explore the impact of product personalization on return rates. This is because if product personalization has a significant effect, the manufacturer can use this inexpensive lever to reduce return rates by allowing customers to personalize products, because adding personalized logos is not costly.

We first discuss the primary variables of interest in §3.2, then the control variables used in the study in §3.3 which include several new measures that control for unobserved consumer heterogeneity.

3.2 Causal Variables of Interest

There are several variables in a production process that can impact output quality of a product, thus returns, as identified in Figure 1. These include worker specialization, workload levels as well as worker ability, worker quality of work, process design, and quality of materials used. Our objective is to examine the potential impact of two important factors, worker specialization and workload levels, on returns while controlling for other characteristics of the production process.

3.2.1 Worker Specialization

Specialization and allocation of tasks have a substantial impact on process performance in terms of typical process measures such as process completion time, process quality, and labor productivity. If workers perform the same task or set of tasks repeatedly over time, worker specialization is higher. Such repetition (higher worker specialization) is likely to result in greater productivity due to learning curve benefits, time saved in switching between different tasks, etc. Learning curves and related benefits of specialization have been studied extensively in the context of productivity (Argote and Epple 1990, Narayanan et al. 2009, Staats and Gino 2012).

However, to the best of our knowledge, the impact of specialization on output quality has been relatively unexplored. For example, Staats and Gino (2012) discuss in great detail the benefits and costs of special-

ization and find evidence in the context of mortgage loan processing operation that worker specialization helps improve short-term worker productivity but workers doing a variety of tasks helps improve longer-term productivity. Still they do not address the issue of quality of output. Some have argued that similar benefits of specialization also extend to quality (Argote and Epple 1990, Compton et al. 1992, Jeang 2015) but such benefits have not been empirically validated. Note that excessive specialization and repetition of the same tasks over long periods of time may also lead to boredom and dissatisfaction resulting in lower productivity (Foley 2008) and quality (Ramdas et al. 2018). Hence, it appears that too much worker specialization can impair productivity and quality. Thus, by linking the literature on the impact of worker specialization on productivity and quality with consumer product returns, we formulate our first hypothesis as below:

Hypothesis 1. The effect of Worker Specialization on the return rate is U-shaped.

The notion of share-based measures of specialization is well known in the economics, healthcare, and OM literature (Peri and Sparber 2009, Sahni et al. 2016, Narayanan et al. 2009, KC and Staats 2012). In the following, we develop two types of share-based measures for *Worker Specialization* which examines the effect of prior experience on the same and different tasks. Since multiple sewers are involved in the processing of an order, we need to measure specialization for the group of workers who processed an order. Let N_i denote the set of sewers that worked on order i . Therefore, the *Worker Specialization* associated with order i should incorporate the specialization of all the workers in N_i .

Our primary measure is similar to Sahni et al. (2016) and is defined as the ratio of the total number of same tasks to the total number of tasks (same + different) during time window t (just prior to the date of working on order i) performed by all the workers belonging to the set N_i . In the computation of this measure, a task performed on order i is considered to be *same* as a previously performed task by a worker if the type of task (one of thirteen types as in Table A1 in the online appendix) and the material code (one of three types) and the product type are identical. Otherwise, they are considered to be different. Mathematically, this can be expressed as $\sum_{j \in N_i} ns_{jt} / (\sum_{j \in N_i} ns_{jt} + \sum_{j \in N_i} nd_{jt})$ where ns_{jt} and nd_{jt} are, respectively, the number of same and different tasks performed by worker j during time window t (we consider several values for t). For example, suppose an order requires two workers and $t = 4$ weeks, and we know from the ERP system that worker 1 (2) performed 10 (30) same tasks and 10 (20) different tasks over the past 4 weeks. Then, the *Worker Specialization* value pertaining to this order is $(10+30)/(10+30+10+20) = 4/7$, which represents an average specialization across the set of workers who process order i . Notice that this measure satisfies three properties: (1) The value of this measure is 0 when no worker has performed the same task(s) in the past. (2) It is equal to 1 when all the workers have performed only the same task(s) in the past and not worked on any other task. (3) It increases as the proportion of the same tasks to total tasks performed by the set of workers for this order increases. Note that the sequence in which tasks were performed, whether same or different, by a worker is not relevant. This is because orders are assigned randomly to workers, as discussed earlier, and so the sequence in which workers perform tasks is random.

Our alternative measure of *Worker Specialization*, which also satisfies the above three properties, is adapted from the Herfindahl-Hirschman Index (HHI) which has a long history in measuring market concentration (Rhoades 1993). Similar measures based on HHI were used in Narayanan et al. (2009), KC and Staats (2012). In particular, we use the following measure: $\sum_{j \in N_i} (\frac{ns_{jt}}{ns_{jt} + nd_{jt}})^2 / |N_i|$ where ns_{jt} and nd_{jt} are as defined above and $|N_i|$ denotes the cardinality of N_i . This measure is similar to the primary measure described earlier except that we square the proportion of same tasks to all tasks for each worker before averaging across the set of workers who process order i .

3.2.2 Workload

Workload levels, which are influenced by the production manager who releases orders to the factory floor, can have a substantial impact on productivity and quality. KC and Terwiesch (2009) cite literature that provides evidence of productivity declining with increasing workload due to fatigue. They also provide evidence that mortality rates increase with increasing workload in cardio-thoracic surgery units. In our context, there is reason to believe that workload levels impact quality of output in sewing. The bins, in which the bundled pieces for various orders in process are placed, represent the work-in-process inventories and are visible to the sewers. Based on prior research, the visible queues are likely to impact the speed at which the sewers work and hence the quality of their work. Unlike the hospital setting in KC and Terwiesch (2009), the workers may be more careless in our setting because their actions do not have “life-and-death” consequences. Thus, we formulate the second hypothesis as follows:

Hypothesis 2. Higher workload leads to higher return rates.

The primary measure we consider for *Workload* is the product type specific total WIP across all stages of production. We consider total WIP because WIP, independent of where it may accumulate, may influence the speed of work of the workers, which in turn impacts quality. For instance, WIP before sewing may impact quality of work of sewers while WIP before packing (after sewing) may impact quality of work in packing/inspection. One unit of WIP is one order. Suppose order i is for product type p . Then, we compute the average work-in-process (WIP_{ip}) in the system for product type p during the period in which order i is processed (during the production lead time of order i). This is a good measure for *Workload* because, as discussed earlier, the WIP is visible to the workers and is likely to influence their speed and quality.

We also consider the WIP at only the sewing stage, which includes WIP in the central bin and all the WIP at individual bins, as part of the robustness tests. We considered this measure because sewing is the most labor-intensive stage of production and sewers are more likely to be influenced by WIP than other workers.

KC and Terwiesch (2009), in a different setting, divide the WIP by the number of resources (beds or transporters) available in the system. In our case, this would be equivalent to dividing by the number of sewers. However, the sewers in our setting may be assigned tasks for different product types based on not only *Workload* levels for each product type but also the task and product types for which they have

the requisite skill. As such, it is not possible to measure the number of sewers by each product type and obtain a *Workload* measure based on average WIP divided by average number of sewers at the product type level. However, we can consider a *Workload* measure at the aggregate level that considers WIP divided by number of sewers across all product types, i.e. $(WIP_i/Sewers_i)$. We consider this alternative measure in the robustness check.

3.2.3 Personalization

The words customization and personalization are often used interchangeably, however there is a subtle difference between these two terms in our setting. Through *customization*, a consumer may choose the fabric type, color, and other products for their specific vehicle accessory. On the other hand, *personalization* takes this idea a step further by offering the consumer an option to add a personalized logo (e.g., a special message of congratulations on one’s birthday, the mascot of one’s alma mater.) The literature has suggested a positive psychological effect of such personalization based on experimental work: a consumer is more likely to be emotionally attached to a personalized product (Mugge et al. 2009) and more likely to request free samples (Howard and Kerin 2004). We study whether such an effect leads to lower consumer returns empirically and formulate our third hypothesis as follows:

Hypothesis 3. Product personalization leads to lower return rate.

We measure *Personalization* of an order using a binary variable (an order requires a logo or not).

3.3 Control Variables

The control variables included in our study can be classified into various categories using the framework in Figure 1, and we discuss each category as follows.

Product, Material, and Order Characteristics. Product type is an important variable that can influence return rates both directly due to fit issues as well as indirectly through quality of output. In a study of an online retailer, Anderson et al. (2009) show variations in returns among different product categories. The extent of fit between a product and a consumer’s preference is likely to be better realized only after purchase for certain products, resulting in differences in return rates across products. Moreover, the importance of fit and uncertainty in fit varies across products (Hess and Mayhew 1997). In our context, there are three major product types as discussed earlier. A slightly larger vehicle cover may be acceptable to a consumer, but a slightly loose seat cover may be regarded as unacceptable and returned. Seat covers cost much more than dash covers, so a consumer is more likely to go through the trouble of returning seat covers. In addition, the quality of output may differ across products due to differences in product type, process design, and process complexity, and this in turn may result in differences in return rates. We also categorize vehicles into twenty-five granular categories (e.g., sedan, coupe, sports utility vehicle) and control for vehicle type since the production process as well as the propensity to return may depend on this variable.

Moreover, a consumer may customize a product by specifying the fabric type. The fabric type may impact the quality of output because the production process is more complex for some fabrics. Also, there are differences in wear and tear among different fabric types. So, within each product type, we control for fabric type using material codes. The company uses three codes—low, medium, and high to categorize materials, which takes into consideration the complexity of processing the fabrics. For example, the material code high includes leather, which is used to make premium seat covers and is more difficult to process due to its thickness and the nature of leather.

While Anderson et al. (2009) do not find that return rates are affected significantly by whether a customer purchased one or multiple items in their order, they suggest that in other applications there may be a significant effect. In this study, about 18.7% of products are sold as part of an order with two or three items. Thus, we utilize the variable *order size* to control for this effect.

Consumer Characteristics. Consumer characteristics are likely to impact return rates as discussed in Anderson et al. (2009) and Su (2008). We do not have data on the individual consumers but we do know their 5-digit zip codes. The absence of data on individual consumer variables is not a limitation in our context because most consumers purchase vehicle accessories infrequently and are likely to have purchased at most once during the study period. We control for several consumer attributes at the zip code level, such as household income and household size, which have been used as control variables in Anderson et al. (2009). In addition, we consider several other control variables at the 5-digit zip code level such as population density (see details in Table A3 in the online appendix). All these variables by 5-digit zip code are obtained from the SimplyAnalytics database (<http://simplyanalytics.com/>), and together they capture a comprehensive set of demographic and socioeconomic attributes of consumers by 5-digit zip code. This has two advantages: First, it controls for various consumer attributes that may influence return rates and thus reduce unobserved heterogeneity. Second, including these control variables helps in eliminating, or at least substantially reducing, endogeneity in estimating the impact of personalization, which will be discussed later in detail in §4.2.3. The comprehensive set of consumer attributes we include control for unobserved consumer characteristics that may be correlated with both the personalization measure and return rate. In addition to the consumer characteristics at the 5-digit zip code level, we also added the 3-digit zip code (first 3 digits) of an order as an additional fixed effect control. This measure controls for factors such as weather and presence/influence of a university, college, or sports team near the consumer’s location. These attributes are likely to be homogeneous within a 3-digit zip code boundary. Return rates may depend upon weather because of weather-related effects on the wear and tear of a product during its warranty period, which in turn impacts the return rate. The presence of a university, college, or sports team may influence the propensity to personalize an order.

Other Production Process Characteristics. The key production process characteristics, *Worker Specialization*, and *Workload* have been discussed earlier in §3.2.1, and §3.2.2 respectively. There are other factors that might influence output quality—sewer ability, and process design—which are discussed next.

Because these characteristics are likely to be correlated with the variables of interest and also influence return rates, we include them as control variables.

Sewer Ability: According to company management, quality of output depends on the ability of sewers.

(a) *Sewer Tenure:* Sewing is a process that requires a lot of skill and it takes about two months for a sewer to achieve the skill level required to achieve an acceptable level of quality and productivity. The sewers are observed closely during their initial two-month probationary period and are asked to leave if they do not achieve acceptable performance. We do not include any orders processed by the sewers during the probationary period. As the sewer acquires more experience, he may get better over time. Therefore, we use employee tenure as a proxy for worker ability, which cannot be observed directly. Since multiple sewers are involved in the production of a product, we use the average tenure of all the sewers who produced a specific order. The average tenure is a good measure if the workers communicate among each other and more experienced workers help the less experienced ones. But the quality of work may depend upon the least experienced worker, especially if communication and cooperation are minimal. While we believe there is communication and cooperation among sewers at the focal firm, we do not know the extent of it, so we also tested an alternative measure as part of the robustness check: the minimum tenure of the sewers producing a specific order. While the order data used in the study is from 2013, we have information on their tenure with the firm from 2007. So, we used the employee's start date prior to 2013 to calculate each sewer's tenure with the company. The other processes such as cutting and packing do not require much skill or training.

(b) *Sewer Quality of Work:* In addition to sewer ability measured by their tenure, we include another control variable to account for a sewer's historical quality of work. In particular, suppose N_i is the set of workers who work on order i . We compute the average return rate of the workers in the set N_i for all orders processed during time window t (just prior to the date of working on order i), for example, $t = 28$. To do this, for each employee in N_i , we measure how many orders she processed in the time window t and how many of those orders were returned, and derive her return rate by dividing the number of returns by the number of orders. Next, we obtain the average return rate of all the workers in N_i . This variable captures the past quality of the work performed by the workers as measured by its impact on return rate. We primarily consider $t = 28$, but we also considered different values of t to account for the quality of work in the past 2 weeks to 3 months as part of our robustness analysis. Additionally, the quality of an order may be determined by the worker with the worst quality performance or equivalently, the worker with the highest return rate among the workers producing an order. So, we also considered the maximum return rate of the set of workers producing an order as part of the robustness checks.

Process Design: The process design for the sewing stage requires a specified number of steps for each product type and a sewer may complete one or more of the steps in the process for an order. So, the number of sewers who work on a specific order can vary. The number of sewers and number of sewing steps in processing an order can have an impact on the quality, especially given that the skill levels of the sewers is not homogeneous. Moreover, it also can influence the tasks performed by the sewer over time and therefore

impact the variable *Worker Specialization*. Therefore, we control for this factor by including a variable called *Process Design*, which is measured as the number of sewers who worked on order i divided by the number of sewing steps specified in the process design for order i . The maximum value of *Process Design* is therefore 100 (unit: %), which represents a scenario wherein one sewer performs exactly one step in the process design. The value of *Process Design* is less than 100 when one sewer performs more than one step in the process. Typically, a sewer completes one step in the process but sometimes the sewer performs two consecutive steps (and occasionally three steps) as we showed in the example in Table 1. Suppose the process design called for four sewing steps and three sewers completed these four steps, then the *Process Design* value is $\frac{3}{4} \times 100 = 75$.

Retailer Characteristics. Based on previous research on the impact of return policies on sales and returns (Wood 2001, Janakiraman et al. 2016 and references therein), the primary retailer attributes that we include are their return policies, which may vary along several dimensions. Below, we discuss the specific dimensions of a retailer’s return policies that are used as control variables. All the measures identified below are proxies for the return effort or cost incurred by the consumer.

Return Period: This measure represents the time window within which a product has to be returned for a refund or replacement for any reason. This window is not applicable for products that are returned because they are found to be defective by the consumer. In that case, the product can be returned at any time during its one-year warranty period but can only be replaced or repaired rather than refunded. In our study, nine retailers impose a 30-day return period, one allows 90 days, and three do not set a maximum return period.

Restocking Fee: Some retailers charge a restocking fee, which is a percentage of the item purchase price, and this fee varies across retailers. A higher restocking fee represents a higher return cost to the consumer. The restocking cost percentage depends only on the retailer and does not vary with product type, material, or personalization.

Return Authorization: Some retailers require a return authorization by calling their customer service or by using their online system before an item can be returned. We have a binary variable to capture whether return authorization is required or not.

Another metric to measure the return policy is whether the retailer allows the consumer to return the product to their brick-and-mortar stores. We do not use this measure because it is highly correlated with the return period in our data set. Specifically, we find that retailers with a lenient return period (90 days and unlimited) are the only ones that allow in-store returns. In the main analysis, for return period we aggregate “30 days” and “90 days” into one category “limited”, and for return authorization we group “required via calling consumer service” and “required via online system” into one category “required”. In the robustness check, we use the uncompressed values for both variables. For further details, please refer to the summary of return policies of the thirteen online retailers in our study in Table A2 in the online appendix.

While return policies are indeed important for explaining return rates, there may be other crucial factors associated with a retailer that influence returns. For example, unlike in experimental studies (e.g., Wood

2001), in our observational study, retailers with different return policies may attract different types of consumers who exhibit different return behaviors. Also, there may be other characteristics of a retailer (e.g., store locations, product lines carried, reputation, brand image) that may influence orders as well as return rates. Thus, we also consider retailer fixed effects as part of the robustness tests presented in §4.4.

Time Fixed Effects. Similar to Anderson et al. (2009) we include fixed effects for both year and month to capture seasonality in return rates as well as any idiosyncratic factors that may have influenced return rates in a given year. The monthly fixed effects capture factors such as consumers purchasing products more impulsively during holiday seasons or a product fading during the summer heat resulting in higher return rates. Time fixed effects also control for occasional free-shipping promotions. Based on our discussion with the manufacturer and retailers in our study, there were no price promotional sales (e.g., 50% off).

4. Return Rate Estimation and Results

In this section, we first discuss the models to estimate the effects of the causal variables of interest and summary statistics in §4.1, then discuss potential endogeneity issues in §4.2, present the estimation results in §4.3, and provide various robustness checks in §4.4.

4.1 Estimation Strategy

The dependent variable is the probability of a return, represented as $\Pr(\text{Return}_i)$ for order i (when there are multiple orders within one order, we treat them separately). We estimate the following model (1) that includes all three causal variables of interest—*Worker Specialization*, *Workload*, and *Personalization*—along with the control variables discussed earlier, represented by the vector \mathbf{X} .

$$\begin{aligned} \text{logit}[\Pr(\text{Return}_i)] = & \mu_0 + \mu_1 \cdot \text{Worker Specialization}_i + \mu_2 \cdot \text{Worker Specialization}_i^2 \\ & + \mu_3 \cdot \text{Workload}_i + \mu_4 \cdot \text{Personalization}_i + \mathbf{X}_i \boldsymbol{\mu}_5 \end{aligned} \quad (1)$$

To test Hypothesis 1 that the effect of *Worker Specialization* on return rates is U-shaped, we consider μ_1 and μ_2 in (1). Specifically, a negative value of μ_1 and a positive value of μ_2 suggest that as *Worker Specialization* increases until $\max\{\frac{-\mu_1}{2\mu_2}, 100\}$, the probability of a return decreases. Afterwards, as *Worker Specialization* continues to increase, the probability of a return starts to increase, which would support Hypothesis 1. This is because a quadratic function $f(x) = ax^2 + bx + c$ given $a > 0, b < 0$ is convex and achieves its minimum at the positive critical point $\frac{-b}{2a}$, and also because *Worker Specialization* does not exceed 100 (unit: %).

To test Hypothesis 2, we use WIP as a measure of *Workload* as we discussed in §3.2.2. A positive value of μ_3 supports hypothesis 2, that is, higher *Workload* leads to higher return rates.

Personalization _{i} is a dummy variable with 1 denoting the presence of a personalized logo. A negative value of μ_4 supports Hypothesis 3, suggesting that personalization of an order leads to a lower return rate.

Summary Statistics

We now provide summary statistics for the causal variables and some key control variables. To reduce the influence of outliers, for the continuous causal and control variables, we first winsorized the bottom and top 2% of the raw data prior to the subsequent analyses (see a discussion on winsorization in §14 in Tukey 1962). The following summary statistics in Table 2 and correlation matrix in Table 3 of key production related variables are based on the winsorized data used for the analyses.

Table 2: Summary Statistics of Production Related Variables

Independent variable	Mean	Std Dev	Min	Max
Worker Specialization (unit: %)				
Team [†] Average (past 4 weeks)	51.9	17.7	10.4	78.6
Team Average (past 3 months)	51.3	17.6	10.9	77.6
Workload (unit: 100 products)				
WIP (all stages)	6.1	4.1	1.3	17.4
WIP (sewing stage)	5.0	1.9	1.3	11.0
Worker Experience and Quality				
Team Average Tenure (unit: month)	62.9	17.2	25.2	95.3
Team Average Return Rate (unit: %, past 4 weeks)	6.5	1.9	2.7	11.3
Team Average Return Rate (unit: %, past 3 months)	6.5	1.6	3.1	9.6
Process Design (unit: %)	95.6	9.4	71.4	100.0

[†] We use the term *team* loosely to denote the set of sewers who worked on the product. There is no fixed team or assembly line as we explained in the manufacturing process in §2.1.

Table 3: Correlation Matrix of Production Related Variables

	1	2	3	4	5	6	7	8
1. Worker Specialization (Team Average, past 4 weeks)	1.00							
2. Worker Specialization (Team Average, past 3 months)	0.97	1.00						
3. Workload (WIP, all stages)	0.03	0.04	1.00					
4. Workload (WIP, sewing stages)	0.08	0.08	0.66	1.00				
5. Team Average Tenure	-0.03	-0.02	-0.37	-0.15	1.00			
6. Team Average Return Rate (past 4 weeks)	0.04	0.03	-0.27	-0.22	0.11	1.00		
7. Team Average Return Rate (past 3 months)	0.04	0.04	-0.29	-0.26	0.13	0.83	1.00	
8. Process Design [†]	0.41	0.41	-0.26	-0.05	0.30	0.22	0.26	1.00

Note: All correlations are significant at the 0.001 level. [‡] See the discussion of variable inflation factor (VIF).

We derived measures of *Worker Specialization* and *Worker Return Rate* for different time windows from $t = 1$ week to 6 months, and then derived the *team* average measures as we discussed in §3.2.1 and §3.3. Note that we use the term *team* loosely to denote the set of sewers who worked on the product. There is no fixed team or assembly line as we explained in the manufacturing process in §2.1. Not surprisingly, there was a high correlation between measures across different t values, for example, the correlation between *Worker Specialization (Team Average, past 4 weeks)* and *Worker Specialization (Team Average, past 3 months)* is 0.97. To be concise, we provide summary statistics for the past 4 weeks and 3 months in Table 2, and use $t = 4$ weeks and $t = 3$ months in the main analyses, and use $t = 1$ week to 6 months as robustness checks.

The measure of *Workload*, the WIP specific to the product type at all stages during the production of an order, had a mean value 6.1 (unit: 100 products), whereas the corresponding sewing stage WIP had a mean value 5.0 which is about 82% of WIP from all stages. Notice the control variable *Process Design* has

a 0.41 correlation with the causal variable *Worker Specialization*, but its small value of $VIF = 1.50$ does not indicate an issue with multicollinearity. To check for the robustness of the results, we run the analyses with and without controlling for this variable. The VIF values for the other variables shown in Table 3 fall in the range between 1.14 and 2.18 and so multicollinearity is not an issue.

The proportion of products with *Personalization* is 2.8%. Details on the proportion of sales by product type, material code, vehicle type, order size, and return period are provided under “Summary statistics of categorical variables” on page A2 in the online appendix.

4.2 Endogeneity Issues

This study includes the main production related covariates that may potentially influence return rates. Moreover, we have included numerous control variables that may impact return rates, as discussed earlier, to minimize potential sources of unobserved heterogeneity. We discuss potential econometric issues with the estimation of the model and steps taken to address them next.

4.2.1 Worker Specialization

As discussed in §2.1 and §3.2.1, work is randomly assigned to workers. This allows us to estimate the impact of *Worker Specialization* on return rates with minimal bias, if any. Recall from our earlier discussion that while sewers are generally assigned to specific product types, the actual tasks they perform for that product type are randomly assigned to them. Moreover, even if sewers are reassigned to a different product type based on workload levels and their ability to perform the tasks, the actual tasks assigned to them are random. However, while the explicit assignment of work is random, we should consider potential omitted variable bias based on implicit factors. For instance, more capable employees may produce better quality and thus achieve lower return rates but their ability is not observable. Such employees may also produce more and hence get assigned more work. However, this does not necessarily induce bias in the estimation of *Worker Specialization* because even if more work is assigned, the assignment of tasks is random. Return rates do not impact *Worker Specialization* either, also because of the random assignment of tasks, thus there is no reverse causality issue.

4.2.2 Workload

Generally speaking, production managers and floor supervisors in manufacturing firms may influence workload levels as well as task assignments, but their influence may not be observable. However, the focal firm in our study operated one shift at the California facility during the period of this study and there was only one production manager and one floor supervisor. Hence, there is no potential bias in our estimation due to the management.

Omitted variable bias: When workload is higher, more capable workers may get assigned more work and thus produce a greater proportion of output. They may also produce better quality which leads to lower return rates. Thus, without controlling for worker ability, the effect of *Workload* is confounded. Hence, we

have to control for the ability of workers to better identify the impact of *Workload*. To do this, we use *Sewer Tenure* and *Sewer Quality of Work* discussed in §3.3 to eliminate potential estimation bias due to sewer ability. In particular, as part of the robustness tests, we control for the *SewerTenure* and *Sewer Quality of Work* of every sewer who works on an order and not just the average values of the team. Finally, we also minimize the potential for omitted variables bias due to process related factors by using the control variable *Process Design* discussed in §3.3.

Reverse causality: Another potential source of estimation bias may be due to reverse causality. Reverse causality can lead to the error term being correlated with the independent variables and can result in biased and inconsistent estimates of the parameters. Can return rates impact the variable *Workload* in our study? Specifically, suppose there are many returns in a particular period. Is it possible that this may impact and thus bias the estimation of the *Workload* on return rates? We think this is unlikely for the following reason. Return rates are only around 8% and only about 5 to 10% of the returned items are reworked or replaced. In fact, we find in the data that rework on average represents less than 0.2% of WIP.

Instrumental variables: Even though estimation bias is not very likely as discussed above, we use two instrumental variables to address potential bias. First, we instrument workload using lagged values of WIP. Recalling that we use WIP as a measure of *Workload*, for any order i , we consider the WIP on the date two weeks prior to the production start date of order i as the lagged value of WIP or *Lagged WIP*. This is a relevant instrument because *Lagged WIP* is likely to be correlated with current period WIP for order i because WIP levels change slowly. The level of WIP depends upon the release of orders into production and the production rate. The production rate cannot be changed easily for the reasons mentioned in §2. This instrument also satisfies the exogeneity condition because the return rate of order i will not be impacted by *Lagged WIP*, except through its effect on current WIP. The second instrumental variable we use is the volume of orders received but not released into production on the production start date of order i . This *Order Volume* is likely to be correlated with WIP, but should not directly impact the return rate of order i except through its impact on WIP. Therefore, it satisfies the relevance and exclusion criteria.

4.2.3 Personalization

Because the decision to order a personalized logo is made by the consumer *before* making the return decision, there is no reverse causality issues due to return rates on the *Personalization* variable.

However, there might be endogeneity in the estimation of *Personalization* because consumers who select to personalize their products may have different characteristics (unknown to the researcher) than those who don't and these may be correlated with the propensity to return a product. Hence, there is potential for selection bias, since we do not know if a personalized order came from consumers who are more likely to personalize or not. To address this endogeneity, we have controlled for a comprehensive set of socio-economic and demographic characteristics as discussed earlier in §3.3. In addition, we explore two different approaches to address this selection bias as discussed next.

Estimation of intent to personalize: We include a proxy variable to estimate and capture this

effect. For each order i , we compute the proportion of orders from the same ZIP3 that were personalized during the period from July 2012 until the week prior to the order date to derive a proxy variable, *Intent To Personalize*. This is a good proxy for the propensity to personalize by consumers as it uses the past history of personalization as a predictor of future intent to personalize (see Bucklin et al. 1998 for a similar approach to capture brand loyalty). The reason we consider ZIP3 rather than ZIP5 level to compute *Intent To Personalize* is that the order volume at ZIP5 level is much smaller with many ZIP Codes having zero or a few orders and so this measure can vary widely (from 0 to 100%) which yields an imprecise measure. For the same reason, we do not consider ZIP3 with less than 8 orders in the study period (bottom 10% at ZIP3 level). This leads to a small reduction in our data size from 179,906 to 172,284.

Instrumental variable: We use an instrumental variable for *Personalization*, *Lagged Sewing Hours*, as explained below. Products that are Personalized typically require more time for sewing, which includes all sewing tasks including embroidery. In our data, the mean time spent in sewing was 1.37 hours for non-personalized products and 1.53 hours for personalized products and the difference is statistically significant. Hence the time spent in sewing (we discuss how we measure this variable and discuss the associated challenges on page A2 in the online appendix), referred to as *Sewing Hours* henceforth, is a relevant instrument for *Personalization*. It is also reasonable to assume that *Sewing Hours* is unrelated to any omitted consumer characteristics that might impact returns. However, orders with longer *Sewing Hours* might lead to a higher quality and thus a lower return rate. For this reason, we consider *Lagged Sewing Hours* measured as the sewing time of the previous order which has the same product and material code and personalization option (binary). We use this new measure as an instrumental variable for each order. It is a relevant instrument because (1) it is correlated with the sewing time of the current order by construction, and (2) the sewing time of an order is correlated with *Personalization*. It meets the exclusion criterion because (1) the sewing time of the previous order should not directly impact the return rate of the current order, and (2) it is not likely to be correlated with any omitted consumer characteristics which may impact the decision to personalize.

Promotion: Based on our discussion with the manufacturer and retailers in our study, there were no price promotional sales (e.g., 50% off) but the firm ran occasional free-shipping promotions which might have slightly influenced the orders received for certain products, and in turn influenced *Worker Specialization* and *Workload* as well as the return rates. We address this issue by controlling for time fixed effects (year and month), retailer fixed effect, and product fixed effects which together help control for the promotional effect.

4.3 Results

The results from estimating equation (1) and variants to it, to test for robustness, are provided in Table 4 and we discuss them in detail next. Models 1 and 2 in Table 4 refer to the estimates of (1) with the only difference between them being that Model 1 uses the three dimensions of retailer return policies—Return Period, Return Authorization, and Restocking fees (described under *Retailer Characteristics* in §3.3)—as control variables while Model 2 uses retailer fixed effects instead.

Table 4: Effect of *Worker Specialization, Workload and Personalization* on Return Rate[†]

	Model 1	Model 2	Model 2a [‡]	Model 2b [‡]	Model 2c [‡]
Hypothesis 1: Worker Specialization					
<i>Team Average</i> (past 4 weeks)	-0.0218 *** (0.0036)	-0.0240 *** (0.0036)		-0.0287 *** (0.0034)	-0.0241 *** (0.0036)
<i>Team Average</i> ² (past 4 weeks)	0.00019 *** (0.00004)	0.00021 *** (0.00004)		0.00026 *** (0.00003)	0.00020 *** (0.00004)
<i>Team Average</i> (past 3 months)			-0.0276 *** (0.0038)		
<i>Team Average</i> ² (past 3 months)			0.00025 *** (0.00004)		
Hypothesis 2: Workload					
<i>WIP</i> (all stages)	0.0219 *** (0.0035)	0.0207 *** (0.0036)	0.0193 *** (0.0036)		0.0250 *** (0.0035)
<i>WIP</i> (sewing stage)				0.0271 *** (0.0064)	
Hypothesis 3: Personalization (yes vs. no)					
	-0.2410 *** (0.0624)	-0.2808 *** (0.0628)	-0.2804 *** (0.0628)	-0.2964 *** (0.0632)	-0.2910 *** (0.0628)
Controls					
Product and Order Characteristics	Included	Included	Included	Included	Included
Product and Material (fixed effect)					
Vehicle Type (fixed effect)					
Order Size (fixed effect)					
Process Design	-0.0113 *** (0.0013)	-0.0106 *** (0.0013)	-0.0107 *** (0.0013)	-0.0116 *** (0.0013)	Not Included
Worker Experience and Quality					
Team Average Tenure	-0.0029 *** (0.0008)	-0.0027 ** (0.0008)	-0.0023 ** (0.0008)	-0.0024 ** (0.0009)	-0.0028 *** (0.0008)
Team Average Return Rate (past 4 weeks)	5.6282 *** (0.7038)	5.6017 *** (0.7044)	5.5649 *** (0.7044)	5.5865 *** (0.7082)	6.0765 *** (0.7036)
Retailer Characteristics					
Return Period (limited vs. unlimited)	-0.6908 *** (0.0346)				
Return Authorization (required vs. not required)	-0.1706 *** (0.0328)				
Restocking Fee%	-0.0134 *** (0.0018)				
Retailer Fixed Effect	Not Included	Included	Included	Included	Included
Consumer Socio-Economic and Demographic Characteristics at ZIP5					
3 Digit Zip Code (ZIP3) (fixed effect)	Included	Included	Included	Included	Included
Seasonality (year and month fixed effect)	Included	Included	Included	Included	Included
Likelihood Ratio (χ^2)	4,923.2 ***	5,061.8 ***	5,070.6 ***	5,028.9 ***	4,991.6 ***
AIC	103,712.6	103,592.1	103,583.3	102,500.9	103,660.2
N (data size)	179,906	179,906	179,906	179,906	179,906

[†] Logit coefficients are provided. Standard errors are shown in parentheses. *** p -value ≤ 0.001 ; ** p -value ≤ 0.01 . To be concise, coefficients of control variables with more than 3 categories are omitted, but available from authors.

[‡] Models 2a, 2b, 2c are variants of Model 2 by changing the time window (from past 4 weeks to 3 months) when measuring *Worker Specialization*, changing *WIP* from including all stages to sewing stage only, and excluding *Process Design* from control variables, respectively.

We consider a model specification with retailer fixed effects because retailers may differ along several dimensions other than return policies and this is captured by retailer fixed effects. We find that the changes between Model 1 and Model 2 in the estimates of μ_1 , μ_2 , μ_3 , and μ_4 are within 0.6 times the standard error. Since the retailer fixed effects control for return policies as well as for other potential differences among retailers, we use retailer fixed effects in most of the remaining model specifications.

We see that $\mu_1 < 0, \mu_2 > 0$ in both models 1 and 2, which together support Hypothesis 1 that the effect of *Worker Specialization* on return rates is U-shaped. When workers perform the same task repeatedly, it results in better quality but too much repetition results in quality degradation. The ratio $\frac{-\mu_1}{2\mu_2}$ in both models 1 and 2 is around 57 (unit: %), suggesting that the probability of return reaches the minimum value when workers perform the same task around 57% of the time. We further find that the effect of *Worker Specialization* is not significantly different when it ranges between 53 and 61. The marginal effect is significantly negative to the left of 53, and positive to the right of 61. For example, when we break down *Worker Specialization* into ten segments each containing the same size of observations, the average marginal effect on the probability of a return is -0.15% in the first segment (from 10.4 to 20.2), and 0.10% in the last segment (from 73.0 to 78.6). We provide the details of the average marginal effects in 10 segments in Table A4 and illustrate them in Figure A1 in the online appendix.

We can see in models 1 and 2 that $\mu_3 > 0$, supporting hypothesis 2. Moreover, $\mu_3 = 0.0207$ in model 2 implies that as *Workload* increases by 1 (unit: 100 orders), the odds ratio of return increases by $e^{\mu_3} - 1 = 2.09\%$. On the other hand, if management reduces WIP by 1 by releasing fewer products into production, the odds ratio of return is reduced by $1 - e^{-\mu_3} = 2.05\%$. We can further compute the average marginal effect of *Workload* on the probability of a return to be 0.2% for each unit increase.

We also find $\mu_4 < 0$ from models 1 and 2 suggesting that product personalization leads to a substantial decrease in the odds ratio of return by $1 - e^{\mu_4} = 21.42\%$ in model 1 and 24.48% in model 2. We find that the average marginal effect of *Personalization* on the probability of a return is -2.4% . So, personalized products are far less likely to be returned.

The return policies of retailers have a substantial impact on return rates as expected and the impact is consistently significant. For example, in Model 1 the limited return period reduces the odds ratio of returns by $1 - e^{-0.6908} = 49.88\%$; when a consumer is required to obtain return authorization, the odds ratio of returns is reduced by $1 - e^{-0.1706} = 15.68\%$; and 1% increase in restocking fees (e.g., 15% to 16%) leads to $1 - e^{-0.0134} = 1.33\%$ reduction in the odds ratio of returns. In terms of average marginal effects on the probability of a return, limited return period, return authorization requirement, and restocking fees are -4.2% , -1.2% , and -0.1% respectively. A limited return period has a stronger effect in reducing returns than *Personalization*, the operational lever under the manufacturer's control. However, we want to point out that many retailers are reluctant to institute strict return policies since they may dampen sales.

In the following, we investigate alternative measures of the causal variables, *Worker Specialization* and *Workload* discussed in §3.2.1, and §3.2.2 respectively.

Different time windows in Worker Specialization: Recall from §3.2.1 that we defined *Worker Specialization* for order i as the ratio of the total number of same tasks to the total number of tasks (same + different) during time window t (prior to the date of working on order i) performed by all the workers that worked on order i . In Model 2 in Table 4, the measure of *Worker Specialization* considers the prior 4 weeks in computing the percentage of tasks that are the same tasks out of all possible tasks performed by the set of workers processing an order. In Model 2a, we consider the prior 3 months (i.e., $t = 3$ months instead of 4 weeks) in the definition of the measure. From the results of Model 2a, it is clear the effect of *Worker Specialization* on the return rate continues to be significant and U-shaped (Hypothesis 1) and the coefficient estimates are similar to those in Model 2. The effect of *Workload* continues to be significant and similar, although a bit weaker. The effect of *Personalization* also remained similar. In fact, we used different t values between 1 week to 4 weeks and 1 month (30 days) to 6 months for the *Worker Specialization* measure. We find small changes in the coefficient of *Worker Specialization* as the time period is varied, but we do not find a statistically significant difference in different time periods (see Table A5 in the online appendix).

Worker Specialization based on HHI: We also explored this alternative measure for *Worker Specialization* for the time windows between the past 1 week and 6 months and continued to find support for Hypothesis 1. As expected, the magnitude of linear and quadratic effects of *Worker Specialization* are different as compared to the primary measure but the signs and significance of coefficients remained the same (see Table A6 in the online appendix). We do not find a statistically significant difference in different time periods. The effect of *Workload* and *Personalization* remained similar.

Alternative measures of Workload: We used an alternative measure of *Workload* wherein we consider only the WIP at the sewing stage. We find that the effect of *Workload* continues to be robust and an increase in *Workload* increases the probability of return as before—see results of Model 2b in Table 4. The coefficient estimates are different primarily because the mean value of WIP at the sewing stage is only 5.0 versus 6.1 for total WIP (see Table 2). However, when we considered another measure, $WIP_i/Sewers_i$, we find that the increase in *Workload* leads to a higher probability of returns, but the coefficient is not significant. We suspect this change is due to the fact that the measure may not accurately capture capacity because (1) as discussed earlier, sewers may be moved across products, (2) the volume that the same sewer can produce can vary substantially across products.

As discussed in §3.3, the variable *Process Design* controls for the number of sewers relative to number of sewing steps in processing an order which can have an impact on the quality. This variable was added to avoid the potential for omitted variable bias. The last column in Table 4, Model 2c, considers a variation of Model 2 which *excludes* the control variable *Process Design*. We find that the coefficient estimates of the variables of interest change only slightly.

4.4 Model Validations and Robustness

In this subsection, we discuss additional robustness checks performed to validate the results on the return rate analysis shown in Models 1a–1b, 2d–2h in Table 5, wherein we only change the control variables.

Table 5: Robustness Checks on Effect of *Worker Specialization*, *Workload* and *Personalization* on Return Rate[†]

	Model 1a	Model 1b [‡]	Model 2d [‡]	Model 2e	Model 2f [¶]	Model 2g	Model 2h
Hypothesis 1: Worker Specialization							
<i>Team Average</i> (past 4 weeks)	-0.0219 *** (0.0036)	-0.0209 *** (0.0037)	-0.0229 *** (0.0037)	-0.0233 *** (0.0036)	-0.0235 *** (0.0039)	-0.0235 *** (0.0036)	-0.0209 *** (0.0037)
<i>Team Average</i> ² (past 4 weeks)	0.00019 *** (0.00004)	0.00019 *** (0.00004)	0.00021 *** (0.00004)	0.00021 *** (0.00004)	0.00021 *** (0.00004)	0.00021 *** (0.00004)	0.00019 *** (0.00004)
Hypothesis 2: Workload							
<i>WIP</i> (all stages)	0.0218 *** (0.0035)	0.0219 *** (0.0036)	0.0206 *** (0.0036)	0.0212 *** (0.0036)	0.0224 *** (0.0038)	0.0205 *** (0.0035)	0.0187 *** (0.0036)
Hypothesis 3: Personalization (yes vs. no)							
	-0.2554 *** (0.0628)	-0.2904 *** (0.0640)	-0.3270 *** (0.0642)	-0.2622 *** (0.0636)	-0.3418 *** (0.0666)	-0.2912 *** (0.0624)	-0.3212 *** (0.0632)
Controls							
Team Average Return Rate (past 4 weeks)	5.5970 *** (0.7041)				5.1021 *** (0.7542)	5.5586 *** (0.6999)	5.3622 *** (0.7061)
Team Average Return Rate (past 3 months)				7.0342 *** (1.0562)			
Team Average Tenure	-0.0029 *** (0.0008)			-0.0029 *** (0.0008)	-0.0026 ** (0.0009)	-0.0026 *** (0.0008)	-0.0026 *** (0.0008)
Each Worker's Return Rate (past 4 weeks)	Not Included	Included	Included	Not Included	Not Included	Not Included	Not Included
Each Worker's Tenure	Not Included	Included	Included	Not Included	Not Included	Not Included	Not Included
3 Digit Zip Code (ZIP3) (fixed effect)	Included	Included	Included	Included	Included	Not Included	Not Included
Vehicle Model (fixed effect)	Not Included	Not Included	Not Included	Not Included	Not Included	Not Included	Included
All Other Controls	Included	Included	Included	Included	Included	Included	Included
Likelihood Ratio (χ^2)	4,934.4 ***	4,933.2 ***	5,072.6 ***	5,043.5 ***	4,590.7 ***	3,797.7 ***	5,216.8 ***
AIC	103,705.5	103,718.7	103,597.3	103,610.4	90,040.4	103,088.2	103,457.1
N	179,906	179,906	179,906	179,906	154,900 ^{¶¶}	179,906	179,906

[†] Logit coefficients are provided. Standard errors are shown in parentheses. *** p -value ≤ 0.001 ; ** p -value ≤ 0.01 .

[‡] Models 1b and 2d control for each worker's return rate and tenure instead of team average return rate and tenure respectively.

[¶] Model 2f excludes orders in the last 6 months of the study, thus the smaller data size.

We expand the return policy measure in the following way. For the variable *Return Period*, we separate the *limited* category into 30 days and 90 days. For the variable *Return Authorization*, we break down the *required* value into “required via calling consumer service” and “required via online system.” We find the corresponding coefficients, reported as Model 1a in Table 5 below, are similar to ones in Model 1 in Table 4.

To control for the historical quality of work of the workers producing an order, we used the *average* return rate of the set of workers producing an order in measuring *Sewer Quality of Work*. But the quality of an order may be determined by the worst worker or equivalently, the maximum return rate among the workers producing an order. So, we considered the maximum return rate, rather than the average, of the set of workers producing an order. To control for the experience of the workers producing an order, we used the *average* tenure of the team of workers that worked on it. But the quality of work of an order may depend upon the least experienced worker and hence the minimum tenure (among the set of workers producing an order) rather than average tenure as discussed earlier. The results obtained by using the alternative measures for *Tenure* and *Sewer Quality of Work* were similar for both Model 1 and 2 (within half standard errors) and are omitted due to space constraints.

In Model 1b and Model 2d shown in Table 5, we further controlled for sewer ability by using the historical quality of work and experience of *each* sewer producing an order, instead of using the team level measures discussed above. The effect of *Personalization* shows up a bit stronger (-0.2904 in Model 1b vs -0.2410 in

Model 1, -0.3270 in Model 2d vs. -0.2808 in Model 2) and is within 0.8 standard errors, whereas the effect of *Worker Specialization* and *Workload* (WIP) are very similar.

We also explored the team average return rate over the previous 3 months for the control variable *Team Average Return Rate*, instead of the 4 weeks used in Model 2 in Table 4 and the results are in Model 2e in Table 5. It is apparent that the coefficient estimates of the variables of interest in Model 2e are similar to those in Model 2.

We also explored if the time window used in the study may have influenced our findings. We considered a variation of Model 2 wherein we dropped the data from the *last* 6 months of the study and the results are reported as Model 2f in Table 5. We find that the effects of *Worker Specialization*, *Workload* and *Personalization* continue to be significant and similar. We also considered a model specification that excludes orders in the first 6 months of the study and a specification wherein we dropped 10% of the orders, randomly selected from the three years, and the results were robust (details available from the authors).

To assess if the 3-digit ZIP Code plays an influential role in the model estimates, we dropped this control variable. Model 2g in Table 5 provides the results of this specification. We find that the coefficients of the variables of interest are about the same as in Model 2. This seems to suggest that consumer and their location characteristics that are captured by 3-digit ZIP Code do not represent potential omitted factors that influence the variables of interest. Additionally, we explored different ways of controlling for consumer characteristics by excluding the population density, using the median age instead of the age groups, using the total household expenditure rather than the expenditures in different categories, either using household expenditure or household income but not both, and found the results to remain robust.

Finally, we included vehicle model as a control variable which contains 896 categories in model 2h in order to understand whether the variation among vehicle models may impact our results. We again find that the coefficients of the variables of interest are about the same as in Model 2. Notice that we do not include 3-digit ZIP Code in this model specification, because our data size is not large enough to accommodate both control variables each containing nearly 900 categories.

Overall, in all the above cases, we find that the coefficients for *Worker Specialization*, *Workload* and *Personalization* are similar across different model specifications and changes in the estimates are small. These results suggest that our results are robust to changes in the measures of the causal variables of interest, time period used to compute the measures, and changes in the control variables.

4.5 Endogeneity Analysis Results

In this subsection, we provide the results of the analyses using instrumental variables and a proxy variable to address potential endogeneity issues discussed in §4.2.

With respect to the variable *WIP*, we first used *Lagged WIP* as an instrumental variable for *Workload* as previously discussed. The Wald weak IV test (Finlay and Magnusson 2009) on *Lagged WIP* rejected the null hypothesis at $p\text{-value} = 0.0002$ and so we conclude that *Lagged WIP* is a strong IV. Using this *Lagged WIP*, we further use the Wald test of exogeneity and failed to reject the exogeneity of causal variable *WIP*

at p -value = 0.6698. The GMM-IVE (Baum et al. 2003) for *WIP* when instrumented by *Lagged WIP* is found to be 0.0209, which is within 0.05 standard error from the logit coefficient estimate of 0.0207 (see Model 2i in Table A7 in online appendix for details). The second instrument we explored for *WIP* is the *Order Volume*. The weak IV hypothesis is rejected at p -value = 0.0028 using the Wald weak IV test. Using *Order Volume* as an IV, we failed to reject the exogeneity of *WIP* at p -value = 0.3402 by the Wald test of exogeneity. The GMM-IVE for *WIP* when instrumented by *Order Volume* is found to be 0.0220 which is within 0.38 standard error from the logit coefficient estimate of 0.0207 (see Model 2j in Table A7 in online appendix for details). The change in the coefficient estimates of the other causal variables are small. Therefore, endogeneity vis-a-vis *Workload* does not appear to be an issue.

Next, we consider endogeneity with respect to the variable *Personalization*. First, we discuss the results of including the proxy variable *Intent To Personalize* in Model 2. We find that it has a coefficient of -0.7904 with standard error 0.3894 and p -value = 0.0424. It seems to suggest that consumers who are more likely to personalize are less likely to return a product at the ZIP3 level. We find that the coefficient estimate of *Personalization* when we include the proxy is within 0.46 standard error of the corresponding coefficient when we don't include the proxy (see Model 2k in Table A7 in online appendix for details). Second, we considered using the IV, *Lagged Sewing Hours* as discussed earlier. The weak IV hypothesis is rejected at p -value = 0.0000 by the Wald weak IV test and indicates that *Lagged Sewing Hours* is a strong IV. Further, using *Lagged Sewing Hours* as an IV we conduct the Wald exogeneity test and find the p -value = 0.1231. The GMM-IVE for *Personalization* when instrumented by *Lagged Sewing Hours* is found to be -0.2668 which is within 0.22 standard error from the logit coefficient estimate of -0.2808 (see Model 2l in Table A7 in online appendix for details). The change in the coefficient estimates of the other causal variables are small. Therefore, endogeneity vis-a-vis *Personalization* does not appear to be an issue.

5. Discussion and Conclusion

The results of our study suggest that *Worker Specialization* is a significant factor that impacts the quality of output, in turn impacting return rates. Moreover, the impact of *Worker Specialization* on return rate is U-shaped. Return rate decreases over a substantial range of *Worker Specialization* values but increases if *Worker Specialization* is too high. As *Worker Specialization* increases, a sewer is able to master a specific task which in turn reduces errors. However, we find that excessive *Worker Specialization*, and in particular, a scenario wherein a sewer performs the same task repeatedly over long periods of time, is not ideal. We find that this is the case whether the *Worker Specialization* is measured over a 4 week period or a 3 month period. In fact, we used different periods (1 week to 6 months) but did not find any statistically significant differences. Our results suggest that it is better for a sewer to periodically perform different tasks, whether these involve working on different products or different materials for the same product or even different set of tasks (e.g. joining or binding or velcro) for the same product and material. Otherwise, boredom or fatigue may set in, resulting in errors. The average level of *Worker Specialization* at the firm in our study is about

52%, which is not far from the optimal level of specialization of 57%. Unlike prior studies which focus on the impact on productivity, we focus on the impact on consumer product returns.

We find that higher *Workload* has a negative impact on return rates. Again, this is consistent with some of the OM literature (KC and Terwiesch 2009, Kim et al. 2017) on the impact of workload on quality, although these studies have primarily focused on health-care settings. In these settings, capacity constraints tend to be hard constraints due to finite bed capacity, while capacity is a soft constraint in our setting because the firm can schedule overtime or extra hours on a Saturday or delay the production of an item. However, these alternatives increase the cost of production or may result in canceled orders, and so there is still pressure on the workers to speed up when *Workload* increases, sacrificing quality. Moreover, the firm cannot increase the number of sewers at short notice. So, the managers face a difficult trade-off between a higher workload which could result in lower quality and a lower workload which may delay orders.

The finding that has the most potential to translate into practice is the utilization of *Personalization* (say, in the form of personalized logos). While the literature has shown that it increases a consumer's attachment to the product, our study appears to be the first one to empirically validate that it has a very strong effect in reducing return rates. Adding logos is not costly, and so it provides the firm with an inexpensive lever to reduce return rates. Also, this does not impact sales negatively, unlike for instance strict manufacturer or retailer return policies, which may discourage returns but simultaneously discourage sales too. Moreover, products with logos have higher margins. In summary, *Personalization* provides multiple benefits.

While the causal variables discussed above are the focus of this study, the effect of the control variables on return rates also provide some interesting insights, especially since our study is the first one to incorporate some of these controls in a study of returns. We focus here only on those control variables whose coefficients are significant and stable and over which the firm has some control.

From a production process perspective, it is interesting to see that *Sewer Tenure* is positively associated with quality. This suggests that worker ability, as proxied by tenure, may be important in reducing defects and return rates. This is recognized by Company *A*, which enjoys low turnover after the initial two-month period, during which they train the employees, and which ensures that they can achieve certain benchmarks in terms of productivity and quality.

We also find that retailer policies have a substantial influence on return rates as one would expect; shorter or limited return periods, higher restocking fees, and requiring an authorization to return all imply lower return rates. Of course, this may be because consumers who are more likely to return products are less likely to purchase in the first place, deterred by the strict return policies. To that extent, such policies are a mixed blessing, because they also reduce sales as has been emphasized in the marketing literature (Hess and Mayhew 1997, Anderson et al. 2009).

It is noteworthy that the production-related factors, *Worker Specialization* and *Workload*, have a significant, albeit small, impact on return rates through their effect on defects despite the fact that there are many possible drivers of returns as discussed in §2. In fact, most of the prior literature has implicitly assumed that

product and retailer attributes are the primary drivers responsible for product returns and do not even discuss production related factors. Our study also shows that manufacturers can offer increased *Personalization* in the form of personalized logos or other product variants as this helps decrease return rates substantially. The impact of product returns is not really as significant for retailers as it is for manufacturers because retailers simply pass along the return to the manufacturer in many instances, as is true in our setting. So, the real cost of returns is borne by the manufacturer and these costs can be substantial.

Acknowledgments

The authors are grateful to Morris A. Cohen, an anonymous associate editor, and three anonymous reviewers for valuable comments that have substantially improved the paper. We are also grateful to Professor S. Siddarth at USC Marshall for his helpful comments on endogeneity. We would also like to thank James Kiselik for his editorial assistance.

References

- Adams S (2017) Optoro Is Building A Billion-Dollar Business Helping Companies Cope With A Glut Of Rejected Stuff, Accessed August 10, 2019, <https://www.forbes.com/sites/susanadams/2017/09/26/optoro-is-building-a-billion-dollar-business-helping-companies-cope-with-a-glut-of-rejected-stuff/>
- Altug MS, Aydinliyim T (2016) Counteracting Strategic Purchase Deferrals The Impact of Online Retailers' Return Policy Decisions. *Manufacturing & Service Operations Management*. 18(3) 376-392.
- Anderson ET, Hansen K, Simester D, Wang LK (2006) How are demand and returns related Theory and empirical evidence. *Kellogg School of Management*.
- Anderson ET, Hansen K, Simester D (2009) The Option Value of Returns Theory and Empirical Evidence. *Marketing Science*. 28(3) 405-423.
- Appriss Retail (2018) 2018 Consumer Returns in the Retail Industry, Accessed August 10, 2019, https://appriss.com/retail/wp-content/uploads/sites/4/2018/12/AR3018_2018-Customer-Returns-in-the-Retail-Industry_Digital.pdf
- Argote L and D. Epple (1990) Learning Curves in Manufacturing. *Science New Series*. 247(4945) 920-924.
- Barnhar B (2019) How to Tap Into Customer Sentiment Analysis & Discover Real Product Insights, Accessed August 10, 2019, <https://www.powerreviews.com/blog/customer-sentiment-analysis/>
- Baum CF, Schaffer ME, Stillman S (2003) Instrumental variables and GMM: Estimation and testing. *The Stata Journal*. 3(1) 1-31.
- Bucklin RE, Gupta S, Siddarth S (1998) Determining Segmentation in Sales Response across Consumer Purchase Behaviors. *Journal of Marketing Research*. 35(2) 189-197.
- Cheng A (2015) Consumers return \$642.6 billion in goods each year, Accessed August 10, 2019, <http://www.marketwatch.com/story/consumers-return-6426-billion-in-goods-each-year-2015-06-18>
- Compton, DW, Dunlap MD and Heim JH (1992) Improving Quality Through the Concept of Learning Curves. in "Manufacturing Systems: Foundations of World-Class Practice" edited by Heim JH and Compton WD. *National Academy Press*. 107-115.
- Dennis S (2018) The Ticking Time Bomb Of E-commerce Returns, Accessed August 10, 2019, <https://www.forbes.com/sites/stevendennis/2018/02/14/the-ticking-time-bomb-of-e-commerce-returns>
- Economist, The (2014) The business of reselling returned shop items What happens to all the goods shoppers don't want, Accessed August 10, 2019, <https://www.economist.com/news/business/21710855-what->

happens-all-goods-shoppers-dont-want-business-reselling-returned-shop-items

- Finlay K, Magnusson LM (2009) Implementing weak-instrument robust tests for a general class of instrumental-variables models. *The Stata Journal*. 9(3) 398-421.
- Fisher ML, Ittner CD (1999) The impact of product variety on automobile assembly operations Empirical evidence and simulation analysis. *Management Science*. 45(6) 771-786.
- Foley DK (2008) Adam's Fallacy A Guide to Economic Theology. *Harvard University Press*. Boston, MA.
- Griffis SE, Rao S, Goldsby TJ, Niranjana TT (2012) The customer consequences of returns in online retailing: An empirical analysis. *Journal of Operations Management*. 30(4) 282-294.
- Guide VDR, Souza GC, Van Wassenhove LNV, Blackburn JD (2006) Time Value of Commercial Product Returns. *Management Science*. 52(8) 1200-1214.
- Hess JD, Mayhew GE (1997) Modeling Merchandise Returns in Direct Marketing. *Journal of Interactive Marketing*. 11(2) 20-35.
- Hong Y, Pavlou PA (2014) Product Fit Uncertainty in Online Markets: Nature, Effects, and Antecedents. *Information Systems Research* 25(2) 328-344.
- Howard DJ, Kerin RA (2004) The Effects of Personalized Product Recommendations on Advertisement Response Rates The "Try This. It Works!" Technique. *Journal of Consumer Psychology*. 14(3) 271-279.
- Howland D (2018) Liquidation startup B-Stock snags \$65M investment, Accessed August 10, 2019, <https://www.retaildive.com/news/liquidation-startup-b-stock-snags-65m-investment/526535/>
- Janakiraman N, Syrdal HA, Freling R (2016) The Effect of Return Policy Leniency on Consumer Purchase and Return Decisions A Meta-analytic Review. *Journal of Retailing*. 92(2) 226-235.
- Jeang A (2015) Learning curves for quality and productivity. *International Journal of Quality and Reliability Management* 32(8) 815-829.
- KC DS, 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 DS, Terwiesch C (2011) The Effects of Focus on Performance Evidence from California Hospitals. *Management Science*. 57(11) 1897-1912.
- KC DS, Staats B (2012) Accumulating a Portfolio of Experience: The Effect of Focal and Related Experience on Surgeon Performance. *Manufacturing & Service Operations Management*. 14(4) 485-699.
- KC DS (2014) Does Multitasking Improve Performance? Evidence from the Emergency Department. *Manufacturing & Service Operations Management*. 16(2) 168-183.
- Kim SH, Pinker E, Rimar J, Bradley E (2017) Refining Workload Measure in Hospital Units: From Census to Acuity-Adjusted Census in Intensive Care Units. *USC Marshall School of Business Working Paper*.
- Lu H and Chen Y. (2012) Returns policy and quality risk in e-business. *Production and Operations Management*. 21(3) 489-503.
- McKinsey & Company (2019) The future of personalization—and how to get ready for it, Accessed August 10, 2019, <https://www.mckinsey.com/business-functions/marketing-and-sales/our-insights/the-future-of-personalization-and-how-to-get-ready-for-it>
- Mollenkopf DA, Rabinovich E, Laseter TM, Boyer KK (2007) Managing internet product returns: A focus on effective service operations. *Decision Sciences*. 38(2) 215-250.
- Mugge R, Schoormans JPL, Schifferstein HNJ (2009) Emotional bonding with personalised products. *Journal of Engineering Design*. 20(5) 467-476.
- Narayanan S, Balasubramanian S, 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.
- Pavlou PA, Liang H, Xue Y (2007) Understanding and mitigating uncertainty in online exchange relationships: A principal-agent perspective. *MIS Quarterly*. 31(1) 105-136.

- Peri G, Sparber C (2009) Task Specialization, Immigration, and Wages. *American Economic Journal: Applied Economics*. 1(3) 135-169.
- Petersen JA, Kumar V (2009) Are Product Returns a Necessary Evil? Antecedents and Consequences. *Journal of Marketing*. 73(3) 35-51.
- Petersen JA, Kumar V (2015) Perceived Risk, Product Returns, and Optimal Resource Allocation Evidence from a Field Experiment. *Journal of Marketing Research*. 52(2) 268-285.
- Pinçe Ç, Ferguson M, Toktay B (2016) Extracting Maximum Value from Consumer Returns Allocating Between Remarketing and Refurbishing for Warranty Claims. *Manufacturing & Service Operations Management*. 18(4) 475-492.
- Powell A, Savin S, Savva N (2012) Physician Workload and Hospital Reimbursement Overworked Physicians Generate Less Revenue per Patient. *Manufacturing & Service Operations Management*. 14(4) 512-528.
- Ramdas K, Saleh K, Stern S, Liu H (2018) Variety and Experience: Learning and Forgetting in the Use of Surgical Devices. *Management Science* 64(6) 2590-2608.
- Rao S, Rabinovich E, Raju D (2004) The role of physical distribution services as determinants of product returns in Internet retailing. *Journal of Operations Management* 32(6) 295-312
- Rhoades SA (1993) The Herfindahl-Herschman index. *Federal Reserve Bulletin*. 79(3).
- Safdar K, Stevens L (2018) Banned From Amazon: The Shoppers Who Make Too Many Returns, Accessed August 10, 2019, <https://www.wsj.com/articles/banned-from-amazon-the-shoppers-who-make-too-many-returns-1526981401>
- Sahni NR, Dalton M, Cutler DM, Birkmeyer JD, Chandra A (2016) Surgeon specialization and operative mortality in United States: retrospective analysis. *BMJ*. 354:i3571
- Shah R, Ball GP, Netessine S (2017) Plant Operations and Product Recalls in the Automotive Industry: An Empirical Investigation. *Management Science*. 63(8) 2397-2771.
- Shang G, Pekgün P, Ferguson M, Galbreth M (2017) How much do online consumers really value free product returns? Evidence from eBay. *Journal of Operations Management*. 53-56:45-62.
- Shulman JD, Coughlan AT, Savaskan RC (2011) Managing Consumer Returns in a Competitive Environment. *Management Science*. 57(2) 347-362.
- Staats BR, Gino F (2012) Specialization and Variety in Repetitive Tasks Evidence from a Japanese Bank. *Management Science*. 58(6) 1141-1159.
- Su X (2008) Consumer Returns Policies and Supply Chain Performance. *Manufacturing & Service Operations Management*. 11(4) 595-612.
- Tan TF, Netessine S (2014) When Does the Devil Make Work? An Empirical Study of the Impact of Workload on Worker Productivity. *Management Science*. 60(6) 1574-1593.
- Taylor FW (1911) *The Principles of Scientific Management*. Harper and Brothers. New York
- Tukey JW (1962) The Future of Data Analysis. *The Annals of Mathematical Statistics*. 33(1) 1-67.
- Wood SL (2001) Remote Purchase Environments The Influence of Return Policy Leniency on Two-Stage Decision Processes. *Journal of Marketing Research*. 38(2) 157-169.

Online Appendix to “Impact of Task-Level Worker Specialization, Workload, and Product Personalization on Consumer Returns”

Hailong Cui^a • Sampath Rajagopalan^b • Amy R. Ward^c

^a*Marshall School of Business, University of Southern California* ^c*Booth School of Business, University of Chicago*

hailongc@marshall.usc.edu • raj@marshall.usc.edu • amy.ward@chicagobooth.edu

August 10, 2019

Table A1: Sewing-Stage Task Name and Description, Applicable Product Types, and Material Codes

Sewing Task Name	Task Description	Products [†]	Materials [‡]
Binding	Add binding at the hem of the fabric to prevent the fabric from tearing.	DC, SC	L, M, H
Elastic	Add elastic on the front and back of cover to provide a snug fit.	VC	L, M
Grommets	Add grommets to the sides of cover for fastening with a lock in windy areas.	VC	L, M
Joining	Joining dozens of pieces of fabric together.	SC	L, M, H
Logo-Embroidery	Add logo via embroidery operation.	DC, SC, VC	L, M, H
Logo-Silkscreen	Add logo via silk screen operation.	DC, SC, VC	L, M, H
Overlock	Perform overlock operation to prevent de-threading of the fabric across seams.	VC	L, M
Overlock Seaming	Assemble all the major panels in cover.	VC	L, M
Pouch	Add the molle patch and pouch at the back to mount tactical bags.	SC	L, M, H
Seaming	Sew the pieces together to form the basic shape of cover.	VC	L, M
Velcro Zipper ABS	Add ABS sheet or Velcro or Zipper at the back of the seat.	SC	L, M, H
Webbing Elastic	Add webbing elastic to secure the SC to the seat of the vehicle.	SC	L, M, H
Zigzag	Using a special machine to create a ZIG ZAG stitch to join all the panels.	DC	L, M

[†] DC = Dash Cover, SC = Seat Cover, VC = Vehicle Cover. [‡] L = Low, M = Medium, H = High.

Table A2: Return Policies of Online Retailers

Retailer	Return Period	Return Authorization	Restocking Fee%	Return to Brick-and-Mortar Store
1	Unlimited	Not Required	0	Allowed
2	30 Days	Required Via Calling Customer Service	15	Not Allowed
3	30 Days	Required Via Online System	0	Not Allowed
4	30 Days	Required Via Online System	20	Not Allowed
5	Unlimited	Not Required	0	Allowed
6	90 Days	Not Required	0	Allowed
7	30 Days	Required Via Online System	15	Not Allowed
8	30 Days	Required Via Online System	15	Not Allowed
9	30 Days	Required Via Online System	15	Not Allowed
10	Unlimited	Required Via Online System	0	Allowed
11	30 Days	Required Via Online System	15	Not Allowed
12	30 Days	Required Via Calling Customer Service	20	Not Allowed
13	30 Days	Required Via Online System	20	Not Allowed

Table A3: Summary Statistics of Consumer Socio-Economic and Demographic Characteristics at Shipping ZIP5

	Mean	Std Dev	Min	Max
Median Household Income (unit: \$1,000)	64.1	21.7	32.5	125.2
Median Household Size	2.8	0.4	2.3	4.0
Population Density (unit: 1,000 people per square mile)	1.8	2.5	0.0	11.3
Percentage of Male Population (unit: %)	49.4	1.6	46.0	55.0
Age Group (unit: %)				
Under 18 Years	23.4	4.4	13.0	33.0
18 to 24 Years	9.3	3.3	5.0	23.0
25 to 34 Years	13.1	3.1	7.0	21.0
35 to 44 Years	12.2	2.0	8.0	17.0
45 to 54 Years	13.6	1.9	9.0	18.0
55 to 64 Years	13.0	2.6	8.0	19.0
65 Years and Older	15.4	5.4	6.0	31.0
Education Attainment (unit: %)				
Less than High School	14.0	8.5	2.0	39.0
High School	27.7	8.2	10.0	45.0
Some College	23.6	4.9	13.0	34.0
Associate's Degree	8.3	2.4	3.0	14.0
Bachelor's Degree	16.8	7.8	5.0	36.0
Post Graduate Degree	9.6	6.1	2.0	29.0
Annual Household Expenditures (unit: \$1,000)				
Total (all categories)	57.0	8.0	28.0	90.4
Apparel and Services	1.8	0.2	1.5	2.3
Entertainment	2.8	0.4	1.8	3.8
Personal Care Products and Services	1.7	0.3	1.1	2.4
Household Furnishings and Equipment	0.7	0.1	0.5	0.9
Vehicle Purchases	3.7	0.5	2.5	4.8

Summary statistics of categorical variables

The breakdown of sales by product type is as follows: dash cover (54.3%), seat cover (38.3%) and vehicle cover (7.4%); and the material codes are low (41.3%), medium (52.7%) and high (6.0%). Any vehicle can be grouped into three broad categories as car (34.5%), truck (61.9%) and van (3.6%) or twenty-five granular categories such as sedan (15.7%), coupe (6.9%) and sports utility vehicles (16.1%) (the full list is available from the authors). Order size equals 1 (81.3%), 2 (17.2%) or 3 (1.5%). The breakdown of sales by retailers with return period is 30 days (51.6%), 90 days (6.4%) and unlimited (42.0%) and further details on the retailer return policies can be found in Table A2 in the online appendix. In the interest of space, we do not provide summary statistics of other categorical variables (e.g., month), but they are available from the authors.

Measuring time spent in sewing

There is a challenge in accurately measuring time spent in sewing. This is because in the focal firm each sewer scans the barcode for an order upon completion of each task, but not at the beginning of each task. For simplicity, consider a product that completes the cutting task 1 at 9:00 a.m., sewing task 1 at 11:00 a.m., sewing task 2 at 1:00 p.m., and packing at 4:00 p.m. One may estimate the sewing hours as 1:00 p.m. (end of sewing) – 9:00 a.m. (end of cutting) = 4 hours. But this measure includes the backlog time between each task, and next we propose an approach to eliminate or at least minimize backlog time included in the

Sewing Hours and thus *Lagged Sewing Hours* measure. In the focal firm in our study, as discussed earlier, management keeps their workers fully utilized (they are rarely idle). Consider an employee who completes sewing task 1 (11:00 a.m.) on order i , and sewing task 2 (11:30 a.m.) on order j on the same day. Because there is almost no idleness, we can assume that sewing task 2 was started at 11:00 a.m. when her task 1 was completed. Thus, we can measure the sewing task 2 time as (11:30 a.m.—11:00 a.m.) = 0.5 hours. Once we have the task time for each employee for each order, we can sum up all the sewing task times for an order to derive the *Sewing Hours* for each product. We do not include non-working hours as part of *Sewing Hours*, such as lunch breaks, evening, weekend or holiday hours when the factory is closed when measuring *Sewing Hours*.

Table A4: Average Marginal Effect vs. Worker Specialization (WS)

WS From	WS To	Average Marginal Effect	WS From	WS To	Average Marginal Effect
10.4	20.2	-0.15%	54.8	58.7	0.03%
20.2	39.7	-0.07%	58.7	62.7	0.04%
39.7	45.9	-0.03%	62.7	67.3	0.06%
45.9	50.5	-0.01%	67.3	73.0	0.08%
50.5	54.8	0.01%	73.0	78.6	0.10%

Figure A1: Average Marginal Effect (vertical) vs. Worker Specialization (horizontal)

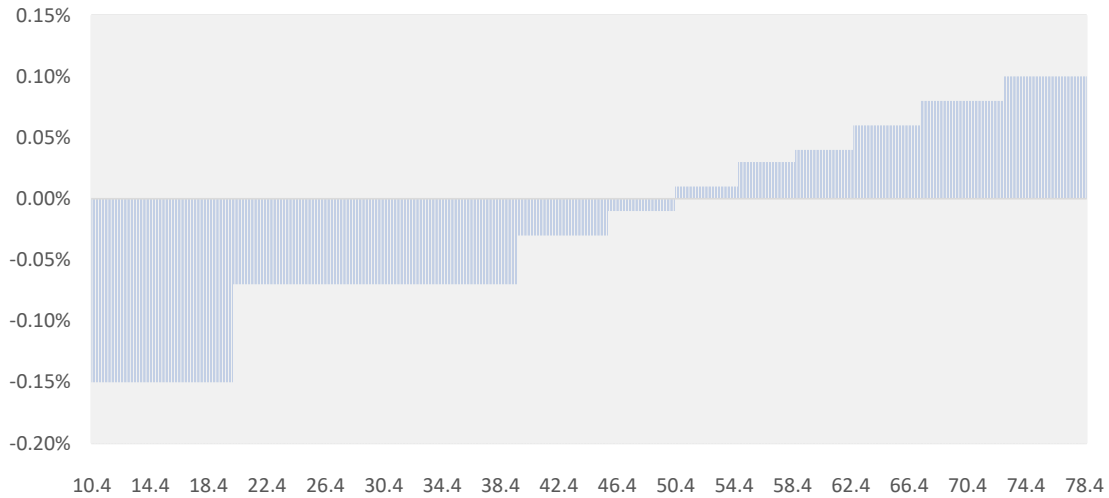


Table A5: Robustness Checks on Effect of *Worker Specialization*, *Workload* and *Personalization* on Return Rate[†] When *Worker Specialization* is Measured During the Past 1 Week to 6 Months (1 month = 30 days)

	Model 1 (past 1 week) [‡]	Model 1 (past 2 weeks)	Model 1 (past 3 weeks)	Model 1 (past 4 weeks)	Model 1 (past 1 month)
Hypothesis 1: Worker Specialization					
<i>Team Average</i>	-0.0200 *** (0.0032)	-0.0222 *** (0.0034)	-0.0209 *** (0.0036)	-0.0218 *** (0.0036)	-0.0216 *** (0.0036)
<i>Team Average</i> ²	0.0002 *** (0.00003)	0.0002 *** (0.00003)	0.0002 *** (0.00004)	0.0002 *** (0.00004)	0.0002 *** (0.00004)
Hypothesis 2: Workload					
<i>WIP</i> (all stages)	0.0228 *** (0.0035)	0.0217 *** (0.0035)	0.0222 *** (0.0035)	0.0219 *** (0.0035)	0.0221 *** (0.0035)
Hypothesis 3: Personalization (yes vs. no)					
	-0.2420 *** (0.0624)	-0.2410 *** (0.0624)	-0.2416 *** (0.0624)	-0.2410 *** (0.0624)	-0.2410 *** (0.0624)
Same controls ^{II}					
	Included	Included	Included	Included	Included
	Model 1 (past 2 months)	Model 1 (past 3 months)	Model 1 (past 4 months)	Model 1 (past 5 months)	Model 1 (past 6 months)
Hypothesis 1: Worker Specialization					
<i>Team Average</i>	-0.0242 *** (0.0037)	-0.0253 *** (0.0038)	-0.0252 *** (0.0039)	-0.0246 *** (0.0039)	-0.0240 *** (0.0039)
<i>Team Average</i> ²	0.0002 *** (0.00004)	0.0002 *** (0.00004)	0.0002 *** (0.00004)	0.0002 *** (0.00004)	0.0002 *** (0.00004)
Hypothesis 2: Workload					
<i>WIP</i> (all stages)	0.0212 *** (0.0035)	0.0205 *** (0.0036)	0.0203 *** (0.0036)	0.0205 *** (0.0036)	0.0208 *** (0.0036)
Hypothesis 3: Personalization (yes vs. no)					
	-0.2398 *** (0.0624)	-0.2408 *** (0.0624)	-0.2424 *** (0.0624)	-0.2440 *** (0.0624)	-0.2452 *** (0.0624)
Same controls					
	Included	Included	Included	Included	Included
	Model 2 (past 1 week)	Model 2 (past 2 weeks)	Model 2 (past 3 weeks)	Model 2 (past 4 weeks)	Model 2 (past 1 month)
Hypothesis 1: Worker Specialization					
<i>Team Average</i>	-0.0217 *** (0.0032)	-0.0243 *** (0.0034)	-0.0231 *** (0.0036)	-0.0240 *** (0.0036)	-0.0239 *** (0.0036)
<i>Team Average</i> ²	0.0002 *** (0.00003)	0.0002 *** (0.00003)	0.0002 *** (0.00004)	0.0002 *** (0.00004)	0.0002 *** (0.00004)
Hypothesis 2: Workload					
<i>WIP</i> (all stages)	0.0219 *** (0.0035)	0.0206 *** (0.0035)	0.0210 *** (0.0036)	0.0207 *** (0.0036)	0.0209 *** (0.0036)
Hypothesis 3: Personalization (yes vs. no)					
	-0.2820 *** (0.0628)	-0.2808 *** (0.0628)	-0.2814 *** (0.0628)	-0.2808 *** (0.0628)	-0.2808 *** (0.0628)
Same controls					
	Included	Included	Included	Included	Included
	Model 2 (past 2 months)	Model 2 (past 3 months)	Model 2 (past 4 months)	Model 2 (past 5 months)	Model 2 (past 6 months)
Hypothesis 1: Worker Specialization					
<i>Team Average</i>	-0.0264 *** (0.0037)	-0.0276 *** (0.0038)	-0.0274 *** (0.0039)	-0.0267 *** (0.0039)	-0.0260 *** (0.0039)
<i>Team Average</i> ²	0.0002 *** (0.00004)	0.0003 *** (0.00004)	0.0002 *** (0.00004)	0.0002 *** (0.00004)	0.0002 *** (0.00004)
Hypothesis 2: Workload					
<i>WIP</i> (all stages)	0.0201 *** (0.0036)	0.0193 *** (0.0036)	0.0191 *** (0.0036)	0.0194 *** (0.0036)	0.0197 *** (0.0036)
Hypothesis 3: Personalization (yes vs. no)					
	-0.2794 *** (0.0628)	-0.2804 *** (0.0628)	-0.2820 *** (0.0628)	-0.2838 *** (0.0628)	-0.2852 *** (0.0628)
Same controls					
	Included	Included	Included	Included	Included

[†] Logit coefficients are provided. Standard errors are shown in parentheses. *** p -value ≤ 0.001 .

[‡] (past t) denotes the past time window t in which *Worker Specialization* is measured.

^{II} See details on controls in Table 4 in the manuscript. Data size is $N = 179,906$ for all model variants shown above.

Table A6: Robustness Checks on Effect of *Worker Specialization*, *Workload* and *Personalization* on Return Rate[†] When *Worker Specialization* is Measured During the Past 1 Week to 6 Months (1 month = 30 days)

	Model 1 (past 1 week) [‡]	Model 1 (past 2 weeks)	Model 1 (past 3 weeks)	Model 1 (past 4 weeks)	Model 1 (past 1 month)
Hypothesis 1: Worker Specialization					
<i>Team Average</i> (based on HHI [⊗])	-0.0128 *** (0.0025)	-0.0141 *** (0.0028)	-0.0144 *** (0.0029)	-0.0157 *** (0.0029)	-0.0152 *** (0.0029)
<i>Team Average</i> ² (based on HHI)	0.0002 *** (0.00004)	0.0002 *** (0.00004)	0.0002 *** (0.00004)	0.0002 *** (0.00004)	0.0002 *** (0.00004)
Hypothesis 2: Workload					
<i>WIP</i> (all stages)	0.0266 *** (0.0033)	0.0261 *** (0.0034)	0.0259 *** (0.0034)	0.0255 *** (0.0034)	0.0257 *** (0.0034)
Hypothesis 3: Personalization (yes vs. no)					
	-0.2518 *** (0.0624)	-0.2536 *** (0.0624)	-0.2532 *** (0.0624)	-0.2532 *** (0.0624)	-0.2530 *** (0.0624)
Same controls [Ⓘ]					
	Included	Included	Included	Included	Included
	Model 1 (past 2 months)	Model 1 (past 3 months)	Model 1 (past 4 months)	Model 1 (past 5 months)	Model 1 (past 6 months)
Hypothesis 1: Worker Specialization					
<i>Team Average</i> (based on HHI)	-0.0146 *** (0.0030)	-0.0142 *** (0.0030)	-0.0136 *** (0.0030)	-0.0133 *** (0.0030)	-0.0127 *** (0.0030)
<i>Team Average</i> ² (based on HHI)	0.0002 *** (0.00004)	0.0002 *** (0.00004)	0.0002 *** (0.00004)	0.0002 *** (0.00005)	0.0002 *** (0.00005)
Hypothesis 2: Workload					
<i>WIP</i> (all stages)	0.0260 *** (0.0034)	0.0261 *** (0.0034)	0.0263 *** (0.0034)	0.0265 *** (0.0034)	0.0268 *** (0.0034)
Hypothesis 3: Personalization (yes vs. no)					
	-0.2524 *** (0.0624)	-0.2528 *** (0.0624)	-0.2542 *** (0.0624)	-0.2560 *** (0.0624)	-0.2574 *** (0.0624)
Same controls					
	Included	Included	Included	Included	Included
	Model 2 (past 1 week)	Model 2 (past 2 weeks)	Model 2 (past 3 weeks)	Model 2 (past 4 weeks)	Model 2 (past 1 month)
Hypothesis 1: Worker Specialization					
<i>Team Average</i> (based on HHI)	-0.0139 *** (0.0025)	-0.0154 *** (0.0028)	-0.0157 *** (0.0028)	-0.0170 *** (0.0029)	-0.0166 *** (0.0029)
<i>Team Average</i> ² (based on HHI)	0.0002 *** (0.00004)	0.0002 *** (0.00004)	0.0002 *** (0.00004)	0.0002 *** (0.00004)	0.0002 *** (0.00004)
Hypothesis 2: Workload					
<i>WIP</i> (all stages)	0.0260 *** (0.0033)	0.0254 *** (0.0034)	0.0252 *** (0.0034)	0.0248 *** (0.0034)	0.0250 *** (0.0034)
Hypothesis 3: Personalization (yes vs. no)					
	-0.2922 *** (0.0628)	-0.2940 *** (0.0628)	-0.2936 *** (0.0628)	-0.2936 *** (0.0628)	-0.2932 *** (0.0628)
Same controls					
	Included	Included	Included	Included	Included
	Model 2 (past 2 months)	Model 2 (past 3 months)	Model 2 (past 4 months)	Model 2 (past 5 months)	Model 2 (past 6 months)
Hypothesis 1: Worker Specialization					
<i>Team Average</i> (based on HHI)	-0.0160 *** (0.0030)	-0.0155 *** (0.0029)	-0.0149 *** (0.0029)	-0.0145 *** (0.0030)	-0.0139 *** (0.0030)
<i>Team Average</i> ² (based on HHI)	0.0002 *** (0.00004)	0.0002 *** (0.00004)	0.0002 *** (0.00004)	0.0002 *** (0.00005)	0.0002 *** (0.00005)
Hypothesis 2: Workload					
<i>WIP</i> (all stages)	0.0253 *** (0.0034)	0.0254 *** (0.0034)	0.0256 *** (0.0034)	0.0259 *** (0.0034)	0.0262 *** (0.0034)
Hypothesis 3: Personalization (yes vs. no)					
	-0.2926 *** (0.0628)	-0.2932 *** (0.0628)	-0.2946 *** (0.0628)	-0.2964 *** (0.0628)	-0.2980 *** (0.0628)
Same controls					
	Included	Included	Included	Included	Included

[†] Logit coefficients are provided. Standard errors are shown in parentheses. *** p -value ≤ 0.001 .

[‡] (past t) denotes the past time window t in which *Worker Specialization* is measured.

[⊗] HHI denotes Herfindahl-Hirschman Index. See details on how this measure is derived in §3.2.1 in the manuscript.

[Ⓘ] See details on controls in Table 4 in the manuscript. Data size is $N = 179,906$ for all model variants shown above.

Table A7: Effect of *Worker Specialization*, *Workload* and *Personalization* on Return Rate[†] using IV

	Model 2 [†]	Model 2i [‡]	Model 2j [‡]	Model 2k [†]	Model 2l [‡]
Hypothesis 1: Worker Specialization					
<i>Team Average</i>	-0.0240 *** (0.0036)	-0.0249 *** (0.0037)	-0.0212 *** (0.0041)	-0.0231 *** (0.0037)	-0.0257 *** (0.0038)
<i>Team Average</i> ²	0.00021 *** (0.00004)	0.00022 *** (0.00004)	0.00018 *** (0.00004)	0.00021 *** (0.00004)	0.00022 *** (0.00004)
Hypothesis 2: Workload					
<i>WIP</i> (all stages)	0.0207 *** (0.0036)	0.0209 *** (0.0045)	0.0220 *** (0.0060)	0.0197 *** (0.0036)	0.0269 *** (0.0037)
Hypothesis 3: Personalization (yes vs. no)					
	-0.2808 *** (0.0628)	-0.2560 *** (0.0607)	-0.2599 *** (0.0610)	-0.3094 *** (0.0644)	-0.2668 *** (0.0663)
Same controls	Included	Included	Included	Included	Included
IntentToPersonalize	Not Included	Not Included	Not Included	Included	Included
<i>N</i> (data size)	179,906	179,906	179,906	172,284	172,284

[†] Logit coefficients are provided in Models 2 and 2k. Standard errors are shown in parentheses. ****p*-value ≤ 0.001 . Model 2k differs from Model 2 in using *Intent To Personalize* as additional control.

[‡] GMM-IVE coefficients are provided in Models 2i, 2j, 2l. Models 2i and 2j instrument *WIP* by *Lagged WIP* and *Order Volume* respectively. Model 2l includes *Intent To Personalize* as additional control, and instruments *Personalization* by *Lagged Sewing Hours*.