Value of High-Quality Logistics: Evidence from a Clash between SF Express and Alibaba

Ruomeng Cui
Goizueta Business School, Emory University, Atlanta, GA 30322, ruomeng.cui@emory.edu

Meng Li
School of Business, Rutgers University, Camden, NJ 08102, meng.li@rutgers.edu

Qiang Li
Department of Supply Chain Management, Rutgers University, Newark NJ, 07102, qiang.li@rutgers.edu

Consumers regard product delivery as an important service component that influences their shopping decisions on online retail platforms. Delivering products to customers in a timely and reliable manner enhances customer experience and companies' profitability. In this research, we explore the extent to which customers value a high-quality delivery experience when shopping online. Our identification strategy exploits a natural experiment: a clash between SF Express and Alibaba, the largest private logistics service provider with the highest reputation in delivery quality in China and the largest online retail platform in China, respectively; the clash resulted in Alibaba unexpectedly removing SF Express as a shipping option from Alibaba's retail platform for 42 hours in June 2017. Using a difference-in-differences design, we analyze the market performance of 129,448 representative stock-keeping units (SKUs) on Alibaba to quantify the economic value of a high-quality delivery service to sales, product variety, and logistics rating. We find that the removal of the high-quality delivery option from Alibaba's retail platform reduced sales by 14.56% during the clash, increased the contribution of long-tail to total sales—sales dispersion—by 3%, but did not impact the variety and logistics rating of sold products. Further, we also identify product characteristics that attenuate the value of high-quality logistics and find that the removal of SF Express is more obstructive for (i) star products as compared to long-tail products, because the same star products are likely to be supplied by competing retail platforms that customers can easily switch to, (ii) expensive products, because customers need a reliable delivery service to protect their valuable items from damages or losses, and (iii) less-discounted products, because customers are more willing to sacrifice the service quality over a price markdown.

Key words: delivery quality, retail platform, sales dispersion, product variety, natural experiment

---

Logistics can spell the difference between success and failure in business. — Heskett (1977).

1. Introduction

Logistics is a critical component in the online retail sphere; it is about getting the product to customers at the right place, time, and cost. In an online purchase, before placing an order, customers evaluate not only the product but also the delivery service. Delivery service includes several important attributes: speed, reliability, and professionalism—that is, whether customers can get the product in a timely and reliable manner and whether customers experience a professional service (Hill 2017). A high-quality delivery service can drive a satisfied customer experience, boost retention, and improve profits. For online retailers, logistics has become a competitive differentiator that drives consumers' shopping decisions. For example, a recent New York Times article reveals that last-minute customers increasingly trust Amazon's delivery service, by observing that Amazon's sales peak about one week prior to Christmas, which is just sufficient time for packages or gifts to be delivered on time (Weise 2018). Thus, numerous companies, from Amazon and Walmart to midsize and small retailers, are debuting and enhancing delivery options to improve service quality (Mottl 2018).

In the traditional view, logistics has often been viewed simply as a cost that needs to be reduced rather than a revenue generator (Lambert and Burduroglu 2000). Consequently, firms have attempted to reduce logistics costs instead of maximizing global profits. However, the extent to which logistics, particularly high-quality logistics, contributes to firms' performance remains unclear. In this research, we aim to show that logistics directly shapes customers' purchasing behavior; it is not merely a backstage cost but rather a prioritized revenue driver that online retailers must strategically and proactively compete over.

In this paper, we explore the value of delivery quality to an online retail platform. In particular, we study how the inclusion or removal of a high-quality delivery option impacts customers' purchasing behavior and satisfaction, and we identify contexts where the delivery quality is important.

Quantifying the value of delivery quality is challenging because it is difficult and often impossible for companies to manipulate delivery options at a company-wide level. For example, in a private interview, the chief strategy officer of SF Express commented, "We believe in the value of logistics quality. But we are struggling with pinpointing the value to retail platforms. [...] We would love to know how much the high quality is worth." In this paper, we adopt a natural experiment approach by exploiting an exogenous shock—altering the logistics quality level on a retail platform—between two business giants, Alibaba Group Holding Limited (Alibaba) and SF Express (SF).

1.1. Research Background and Natural Experiment

The e-commerce market in China has been growing rapidly in the recent decade, with 772 million internet users and a gross merchandise volume (GMV) of ¥6.1 trillion ($878 billion) in 2017 (Fbicgroup 2018, Statista 2018b).[1] In China, Alibaba is the top retail e-commerce platform, with 58.2% market share, followed by JD.com (JD) with 16.3% share (eMarketer 2018, Statista 2018a). Chinese customers are becoming increasingly sensitive to delivery service (Joerss et al. 2016). The

[1] In this paper, ¥ is the currency symbol for Chinese yuan.

speed and efficiency of delivery increased by 14%, and the volume of e-commerce packages also increased by 40% in 2015 (Meola 2016).

Alibaba is a multinational e-commerce, retail, Internet, and technology conglomerate that operates the largest online sales marketplace in China; Alibaba achieved a GMV of ¥3.8 trillion ($547 billion) in 2017, which is three times as large as its US counterpart Amazon's GMV, and aims to attain a GMV of ¥6 trillion by 2020 (Alibaba 2017, Lee 2016, Thomas and Reagan 2018). The company offers a wide selection of products to customers, with over 800 million SKUs; in comparison, Amazon offers 560 million SKUs (Alibaba 2018, Culey 2018).

On Alibaba's retail platform, 99% of sales originate from over 11 million third-party merchants (Talmassons 2016). Merchants have two ways of managing their logistics: (i) self-manage inventory and use third-party delivery carriers to ship products to customers, or (ii) delegate the entire logistics service to Alibaba. In 2013, Alibaba founded Cainiao Smart Logistics Network Limited (Cainiao) to handle inventory and parcel deliveries and opened up this service to all merchants on Alibaba. If a merchant opts to use Cainiao's service, Cainiao handles all its deliveries. If a merchant opts to not use Cainiao, the merchant needs to offer third-party carrier options to customers. In this paper, we call merchants that use Cainiao's service *Cainiao merchants* and merchants that self-manage logistics *non-Cainiao merchants*.

Delivery provider options selected by the merchants are displayed on a merchant's product page. Specifically, Cainiao merchants display only Cainiao delivery on their pages, whereas non-Cainiao merchants display the assigned carrier options (other than Cainiao). When buying from non-Cainiao merchants, if not satisfied with the listing, a customer can contact the merchant and request to use any other unlisted delivery company. In most cases, the merchant will satisfy the customer's request. Occasionally, the customer may need a high-quality delivery for a specific order and may request a premium carrier.[2]

SF Express is one of the third-party carrier providers on Alibaba. It is China's largest private carrier by market value (Wang and Glamann 2018). In 2017, SF handled a total of 3.05 billion shipments, an 18% increase compared to 2016. SF is best known for its high-quality delivery service and excellent customer service. All its shipments are distributed by its full-time employees to customers' doorsteps via self-operated (proprietary) trucks and airplanes, or high-speed trains, which helps SF achieve speedy, reliable, and professional pickup and delivery services (Zhu and Lane 2015).[3] For example, SF is 50% faster than public delivery companies and 20% faster than

[2] In this case, the merchant may or may not charge the customer an extra fee to cover additional costs, depending on the type of contract signed with the premium carrier.

[3] In 2017, SF charges ¥23 ($3) per package, on average, and the industry average is ¥12.38 ($1.8) (Leadingir 2018). SF charges a higher price because it offers the best quality in service.

private delivery companies in China (Liu and Kang 2015). From consumers' perspective, SF is also perceived as the best delivery provider. In fact, SF has ranked first in terms of customer satisfaction surveys (over dimensions including speed, reliability, pickup, drop-off, customer service, and after-sales) conducted by the State Post Bureau of China for over eight consecutive years from 2009 to 2017 (Mcmillan 2018, State Post Bureau 2017b). Figure 3 in the appendix summarizes major carriers' consumer complaint rates in terms of delay, damaged and lost, and delivery service surveyed by the State Post Bureau of China in December, 2017; SF has a much lower complaint rate than others, including the second-place carrier, which shows SF's superiority in service quality.

Alibaba and SF have been close partners over the years. A significant proportion of deliveries on Alibaba were shipped via SF (Zhu and Lane 2015, Soo 2017), and one-fifth of SF's parcel orders are from Alibaba (Bloomberg News 2017). However, a clash between the two suddenly began on June 1, 2017, although "it is unclear what originally triggered their spat" (Bloomberg News 2017).[4] SF unexpectedly blocked Cainiao's access to its parcel data. Within several hours, Alibaba promptly removed SF as a shipping option for its merchants and customers:

> *We are surprised and disappointed by SF's abrupt action to stop providing the information that is necessary for the smooth completion of parcel deliveries. To protect more than a million of consumers and merchants from potential parcel losses, we have no option but to remove SF as a delivery option on Cainiao's network.*
>
> — Cainiao spokeswoman, as quoted in Bloomberg News (2017) and Ye (2017).

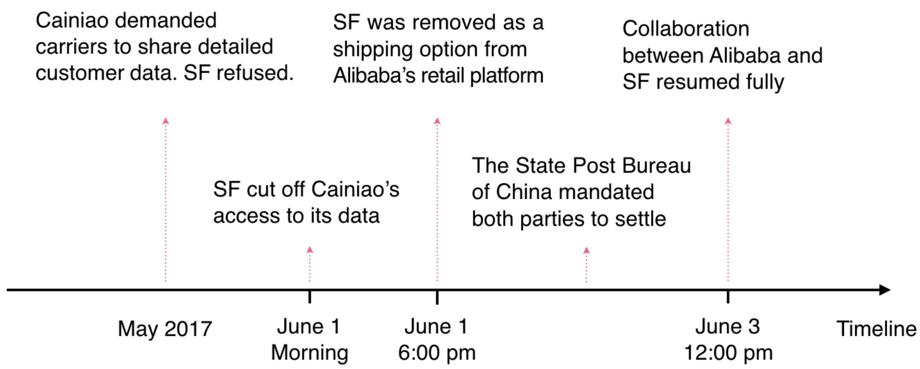
After SF was removed from Alibaba's retail platform at around 6:00 p.m. (Chinanews 2017), merchants were not able to use SF's service and, consequently, customers were unable to choose SF on Alibaba to deliver their orders.[5] Then, the State Post Bureau of China, which regulates the logistics industry, stepped in and asked both parties to reach a settlement for the sake of consumer interest and market order. Finally, cooperation between Alibaba and SF fully resumed at 12:00 p.m. on June 3, 2017 (Soo 2017, State Post Bureau 2017a). Figure 1 summarizes the timeline of the clash.

This clash serves as an ideal setting for a natural experiment to identify the value of high-quality delivery on a retail platform and how it drives consumers' buying decisions. First, the shock was exogenous—the sudden access cutoff was unexpected by Alibaba, and the corresponding removal of SF as a shipping option was an unexpected move for SF, merchants, and consumers;

[4] Their dispute may have arisen over a desire to control the customer information generated by vast orders on Alibaba. Cainiao has been demanding parcel carriers to share their detailed customer data in order to optimize its own delivery routes. However, SF did not share the data demanded by Cainiao because it deemed the request irrelevant (Bloomberg News 2017).

[5] According to Bloomberg News (2017) and Bai (2017), SF was removed as a shipping option from Alibaba's system so that merchants could not use SF as a delivery option. Therefore, any order that requested SF could not be placed.

Figure 1 Timeline of the Clash



therefore, product sales were barely exposed to unobserved confounders, such as planned promotion or search algorithm with regard to affected products, as confirmed by our interviews with company executives. The resumption was exogenous as well—SF and Alibaba were mandated to resolve the differences by the Chinese government. Second, as the delivery company with the highest standards and reputation in delivery quality, SF's removal led to a direct reduction in delivery quality level and consumers' perception of logistics quality on Alibaba's platform; the resumption of cooperation between the two led to an improvement in consumers' perception and the actual logistics quality level. Specifically, based on the back-of-envelope calculation using our data set, SF is 10 hours faster, on average, than the other carriers on Alibaba.[6] This allows us to investigate the value (cost) for a platform to (not) have a high-quality delivery service. Third, the shock did not affect all products—Cainiao merchants were not affected because they had only the Cainiao delivery option, whereas non-Cainiao merchants were affected as these customers could choose from among various carriers, including SF.[7] Fourth, the shock was short—it lasted for 42 hours, which significantly reduced the time and possibility for Alibaba to adjust its searching, pricing, or promotion plans, and for merchants to adjust logistics strategies. Last, the clash between the two business giants—that involves vast product categories, millions of consumers, merchants, and billions of transactions—represents a rich and clean research context. The findings drawn from this large-scale study can guide logistics strategies for merchants and retail platforms.

[6] We first use a difference-in-differences analysis to compute how much slower the average delivery speed became during the clash. Then, we divide this by SF's delivery share of Alibaba's total orders, which is obtained through private interviews with Alibaba and SF.

[7] Customers could no longer find SF listed as a shipping option for non-Cainiao merchants. We interviewed over 10 merchants on Alibaba. Many non-Cainiao merchants remembered that when the clash happened, they were busy explaining to customers why they could not choose SF and apologizing for the inconvenience. On the other hand, Cainiao merchants can only choose Cainiao delivery so their operations were not impacted.

1.2. Findings

We obtained 14,438,123 transactions from 129,448 representative SKUs on Alibaba around the time of the clash. The data includes standard product information—i.e., price and product category—and how satisfied customers felt with the delivery of their orders—i.e., logistics rating. We measure three outcome variables: sales, the variety of sold products, and customer satisfaction with the logistics. In addition, we measure variations in the explanatory variable—the logistics quality level—introduced by the natural experiment.

We use the difference-in-differences (DID) design that exploits the natural shock in logistics quality, comparing sales, product variety, and the logistics rating of the treated products (products that are sold by non-Cainiao merchants) and untreated products (products that are sold by Cainiao merchants). We verify the validity of our identification by showing that the treated and control products followed the same trend before the treatment.

We find that the (unavailability) availability of a high-quality delivery service significantly (decreases) improves the platform's performance. In particular, removing the highest-quality delivery option reduced sales by 14.56%—which is 7,501.14 fewer units sold per hour for the platform, and 20.42 fewer units sold per hour for merchants; reinstating the highest-quality option resumed the sales volume. The lost sales are likely due to consumers' switching to other platforms. Based on the search data of the leading search engine, we find that the search volume of "JD logistics"—JD being the second largest online retailer in China with the second best delivery quality after SF—increased by 23.36% during the clash. This indicates that customers who are sensitive to delivery quality ceased to make a purchase on Alibaba and switched to purchase from a competing retailer.

We further investigate how the value of high-quality logistics is contingent on product characteristics. First, we investigate two types of products: "star products," which are popular products amounting to 80% of total sales, and "long-tail products," which are niche products amounting to 20% of total sales. We find that high-quality logistics is more beneficial for star products than long-tail products: the clash increased sales dispersion—the contribution of long-tail to total sales—by 3%. Because it is likely that popular products are also carried and sold by competing platforms, customers who are not satisfied with the logistics quality can easily switch to another platform that offers high-quality logistics. We find further evidence that supports our conjecture: the width of product variety—the number of unique SKUs sold per hour—was not affected by the shock. This implies that customers of long-tail products, even if not satisfied with the delivery, have no choice but to remain on the current platform. Second, we show that expensive products are more adversely affected by a lower logistics quality. In other words, customers purchasing high-value products care more about delivery quality. Intuitively, customers demand a reliable delivery service to protect their valuable purchases from damages or losses. Last, we find that a lower logistics quality is more harmful when a product is less discounted. We conjecture that the price markdown and the logistics quality are substitutes, that is, a retail platform compensates customers for mediocre logistics by offering a deep discount.

Last, we find that the logistics rating of products was not affected by the treatment. Quality-sensitive customers would choose not to purchase on the current platform, whereas customers who are less sensitive to delivery quality would continue to shop on the platform even in the absence of a high-quality delivery service. The remaining customers are the ones that rate the delivery service. Because they are not very selective about logistics, their evaluations tend to be the same as before.

Further, we supplement our DID analysis with matching. We identify robust estimations using thousands of matched product pairs in which a pair has similar price, product category, and hourly sales in the pre-treatment period but receives different treatments, thereby showing that our results are highly consistent. As a robustness check, we examine and eliminate potential confounding factors, such as cannibalization, promotion, Alibaba's strategic reactions to the clash, and news effect.

2. Related Literature

Our work is related to the literature on logistics quality, especially in terms of speed and reliability. The literature relies on a key theoretical assumption that a higher delivery quality results in higher consumer demand and satisfaction (Kumar et al. 1997). Consequently, online retailers have identified quick delivery as a critical platform feature (Brynjolfsson et al. 2009). The literature lacks empirical evidence on whether and the extent to which delivery quality matters to customers and shapes their shopping decisions; existing evidence is still largely anecdotal rather than quantitative. Recent works empirically quantify the value of one dimension of delivery quality—speed. In particular, Fisher et al. (2018) empirically study the value of an expedited delivery speed on demand; Calvo et al. (2018) and Cui and Sun (2018) investigate how to promise customers delivery speed. Our research complements the literature by studying a broader spectrum of delivery quality, including both speed and reliability. Our findings indicate the importance of strategically managing the delivery quality on a retail platform.

In addition, our paper contributes to the growing literature on how to manage platforms. The existing literature has investigated how platforms manage entry (Zhu and Iansiti 2012), information (Wagner et al. 2019), product variety (Cottrell and Nault 2004), consumer satisfaction (Chevalier and Mayzlin 2006), and competition (Zhu and Liu 2018). We complement this stream by showing that logistics quality—an important yet unexplored component—can give online retail platforms a competitive edge in attracting and retaining customers.

This paper sheds light on a key market outcome for retail platforms—product variety (Ramdas 2003). Traditional brick-and-mortar stores often focus on star (popular) products because of their high turnover and lower inventory costs. On the other hand, online platforms offer merchants an opportunity to sell long-tail products for niche markets that would not be financially viable in traditional markets (Brynjolfsson et al. 2010). While the literature has shown the importance of product variety or sales dispersion in inventory management (Gallino et al. 2016), consumer purchase (Ton and Raman 2010, Cachon et al. 2018), manufacturing processes (Ramdas et al. 2003), and demand concentration (Tan et al. 2017), the relationship between product variety, sales dispersion, and delivery quality has not been directly investigated. We contribute to this literature by studying the synergy between delivery quality on the one hand and product variety and sales dispersion on the other. We particularly find that the long-tail products are less adversely affected by a low delivery quality as compared to star products possibly due to a less competitive environment, which highlights the advantage of offering long-tails on a retail platform. This result echoes Porter's differentiation strategy—differentiating products shields companies from direct competition.

Overall, our work contributes to the general realm of operations strategy. This realm of literature links a firm's performance with its operations strategy, including the pattern of structural (e.g., service design, supply chain design, and performance measurement) and infrastructural choices (e.g., warehousing, plant, and facility network). For example, Roth and Jackson III (1995) investigate the impact of operations capabilities on a service firm's strategic behavior. Randall et al. (2006) show that a retailer decision to keep fulfillment capabilities in-house should fit with its competitive priorities. Ramdas and Randall (2008) examine how component sharing strategy impacts product quality and reliability. We follow suit to empirically examine a key operations strategy—logistics—that affects both structural and infrastructural components because it involves choices of delivery options, process technology, and warehousing. Thus, our work helps guide decisions regarding logistics strategies that will support firms' objectives.

3. Theoretical Motivation

Next, we theorize customers' responses to different levels of delivery quality driven by the clash between Alibaba and SF. We explore two effects: (i) the overall effect of delivery quality, and (ii) the relative effect of delivery quality.

3.1. Value of Delivery Quality

When deciding whether or not to buy a product, apart from evaluating the product itself, a customer also assesses the delivery service—how fast and how reliably the available carriers can deliver the product. On Alibaba's retail platform, the customer can choose from a list of delivery companies selected by the merchant. Based on past experience and the reputation of delivery providers, the customer selects a specific one that satisfies her expectation of delivery quality. If the desired carrier is not on the list, the customer can ask the merchant to accommodate her preference and, in most cases, this request is granted. If the desired carrier is still not available, the customer will either (i) complete the purchase using available carriers if delivery quality is not her top concern, or else (ii) purchase the product from another retail platform that is served by the requested carrier or carriers of similar quality.

Sales. Recall that SF is the delivery courier that provides the highest-quality delivery service in the market. Compared with its competitors, SF has a competitive edge in delivery speed, delivery reliability (the ability to deliver on or before the promised due date), professional pickup and drop-offs, and other related services. Consequently, when SF was removed as a shipping option from Alibaba's entire platform, prospective consumers who wished to use SF may have simply switched to another retail platform. This case is particularly likely whenever consumers can find the same product from rival platforms—for example JD, whose integrated logistics system ranks immediately after SF in customer satisfaction, according to the State Post Bureau of China. Therefore, we hypothesize that a decline in service quality reduces sales, and a reinstatement of service quality restores sales volume to normal.

HYPOTHESIS 1. *The removal of the high-quality delivery option reduces sales, whereas the reinstatement of the high-quality delivery option restores sales.*

Product variety and sales dispersion. Online retail platforms often adopt the long-tail strategy to manage product variety—selling a large number of unique items in relatively small quantities (the "long tail") in addition to selling fewer popular items in large quantities (the "head").[8] The popular products at the "head" of sales distribution are likely carried and sold by competing platforms; consequently, customers who are not satisfied with the logistics can switch platforms to buy the product. Long-tail products, which usually account for over 80% of SKUs, determine the width of product variety. They aim for more differentiated niche markets, which face lower competition (Brynjolfsson et al. 2010). Thus, long-tail product customers have few alternative platforms to switch to, even if they are not satisfied with the logistics. This is particularly likely, given that Alibaba's main competitor JD has only 4% of Aliabab's product variety (Hong 2017, Alibaba 2018). Consequently, although the sales of popular products decline due to lower delivery quality, long-tail products—which determine the width of product variety—are unlikely to be affected. Therefore, we expect that the change in delivery quality does not significantly affect product variety. However,

[8] We follow the literature's definition of product variety, which distinguishes product variants with features such as colors and flavors at the SKU level.

the sales dispersion, which is defined as the contribution of long-tail to total sales, will increase; see Figure 4 in the appendix for an illustration.

HYPOTHESIS 2. *The removal and subsequent reinstatement of the high-quality delivery option do not significantly affect product variety, whereas the removal of the high-quality delivery option increases sales dispersion and the reinstatement of the high-quality delivery option restores the original level of sales dispersion.*

Logistics rating. Before placing an order online, a consumer forms an expectation of delivery quality. After receiving her order, the customer can rate her satisfaction with the delivery service. In the absence of high-quality logistics, customers who shop on the platform anyway are the ones that provide the ratings. These remaining customers are less sensitive to delivery quality: they do not sense or care about the difference in delivery quality. Thus, they are likely to feel as satisfied with the logistics, as earlier.[9] However, a few remaining long-tail product customers might demand high-quality delivery because they have no choice but to remain on the current platform. Although they might not be as satisfied, their purchase volume accounts for a small proportion of total sales and, thus, is less likely to influence the platform's overall logistics rating. Therefore, we conjecture that the logistics rating remains unchanged when the highest-quality delivery service was unavailable.

HYPOTHESIS 3. *The removal and subsequent reinstatement of the high-quality delivery option do not significantly affect logistics rating.*

3.2. Moderating Factors

Next, we study contexts where the delivery quality matters most. We particularly examine three product-specific characteristics—popularity, price, and price markdown—that moderate the value of delivery quality in sales.

Product popularity. As discussed above, popular products (at the head of sales distribution) usually face more intense competition from rival retail platforms. Their customers can easily switch platforms if not satisfied with the logistics service. Less popular products (at the tail of sales distribution) face milder competition and, thus, their customers, although reluctantly, might still purchase on the platform. Therefore, we hypothesize that popular products are more adversely affected by delivery quality.

HYPOTHESIS 4. *A higher product popularity amplifies the value of high-quality delivery.*

[9] Prior to the quality change, quality-sensitive customers tend to have a higher expectation and, thus, may demand a high-quality delivery option; on the other hand, quality-insensitive customers tend to have a lower expectation and thus, may not demand a high-quality delivery. Consequently, their satisfaction rates, which equal the difference between experienced quality and the expected quality, can be similar. Because the customers who left and those that stayed originally had a similar level of satisfaction and the remaining customers rate the same as before, the overall logistics rating is not affected.

Product price. Consumers' demand for delivery quality differs according to whether they are buying cheap or expensive products. This is because a customer considers the possible size of the loss that she would incur (especially if the product is expensive) should the delivery service fail to complete the order. Intuitively, a more expensive item would necessitate a more reliable carrier to protect it from being damaged or lost. Therefore, customers who buy expensive products are the first to react to the absence of high-quality logistics. Then, we speculate that a higher product price amplifies the value of delivery quality.

HYPOTHESIS 5. *A higher product price amplifies the value of the high-quality delivery.*

Price markdown. Retail platforms often run promotions to attract and retain customers. Customers trade off the discount and service quality. When benefiting from a deeper discount, customers are more likely to tolerate a mediocre delivery. In other words, the price markdown and the delivery quality are substitutes: customers are more willing to sacrifice delivery quality for a higher price markdown.

HYPOTHESIS 6. *A higher price markdown weakens the value of high-quality delivery.*

4. Data

We obtain sales- and logistics-related data on Alibaba's platform from the 2018 MSOM data-driven research challenge. This data includes details on customer orders and order fulfillment logistics from Jan 1, 2017 to August 31, 2017. "All data are anonymized [...]. The structure of the data is representative of the overall business operation, and [...] insights based on mining on this data would be beneficial to the industry as a whole" (Tianchi 2018). We obtain two types of data: market outcome data and product characteristics data.

Transaction data. The data records detailed information associated with each transaction. Each transaction record includes the time of the order, merchant ID, customer ID, products (SKUs), and their prices at the time of the purchase. For each merchant, we observe whether it is a non-Cainiao merchant or a Cainiao merchant. We exclude observations of digital products such as software, internet coupons, serial codes for games, and gift cards, which account for 0.5% of total sales.[10]

To obtain the measure in sales, we first aggregate sales at the hourly level. We then aggregate hourly sales at the group, merchant, and product levels, respectively. Note that some merchants use Cainiao fulfillment to manage certain products, while self-managing the rest of products. We compute the outcome variables for these merchants separately for the treatment and control groups. Because one product has only one fulfillment method throughout our sample, there is no double

[10] We define products that were never delivered by a logistics company from January to May in 2017 as digital goods.

counting in our aggregation process. To obtain the measure in product variety, we count the number of unique SKUs sold per hour, which measures the spectrum of product lines.

For each transaction, we also observe the customer's logistics rating. After receiving the order, the customer can score her satisfaction with the delivery service from 1 to 5—a higher score means higher satisfaction. All the products within an order receive the same score. If the customer does not review the logistics, the score is missing. For each product, we measure its logistics rating by averaging its logistics scores within a specific time period.

Product data. We also collect data on product characteristics. We observe the category of each product, which we use to control for category attributes. There are 397 categories in our sample. To measure product popularity, we identify star products at the head of sales distribution, and long-tail products at the tail of the distribution. We define star products as the ones that account for 80% of total sales from January to May in 2017 and long-tail products as the remainder that account for 20% of total sales. A product's price can change over time. To obtain the average price of a product, we average its price within a specific time period. Based on the product price, we derive the measure of discount. We first compute the baseline price for each product over January 1 to May 31, 2017, that is, before the shock occurred. We then compute the discount as the current price relative to the baseline price.

Descriptive statistics. Table 1 presents the summary statistics at the product or hourly level during the historical period prior to the shock—from January 1 to May 31, 2017—and during the natural experiment period—from June 1 to June 3, 2017. Over the historical period that covers 4,344 hours, the sample includes 233,318 SKUs. The platform sold 8,601 unique SKUs and 51,503.58 units of products per hour, of which 40,156.64 were star products. The average price across products was ¥231.51, and the average logistics rating was 4.81. Over the natural experiment period of 42 hours, the sample includes 1,278,433 transactions and 129,448 SKUs, which is less than the SKU number in the historical period. In this period, the platform sold 8,958 SKUs and 54,040.02 units of products per hour, of which 39,149.62 were star products. Relative to the historical period, the average product price fell to ¥167.63, and the average logistics rating remained almost the same at 4.82.

5. Identification Strategy

We exploit a natural experiment setting, which has the advantage of providing a high external validity on causal inferences (Cook et al. 2002). In particular, we adopt a DID approach to identify the value of a high-quality logistics option on retail platforms.

Table 1 Summary Statistics for Alibaba Products from January 1 to June 3, 2017.

Time window	Level	Variable	Mean	Std. dev	Min	Max	N
Historical period (Jan. 1–May 31) 233,318 SKUs	Hour	Total sales	51,503.58	42,864.53	373	961,836	4,344
		Total "star" sales	40,156.64	33,589.34	270	567,060	4,344
		Product variety	8,601	4,716	299	36,365	4,344
	Product	Price	231.51	929.78	0.01	80,579.31	233,318
		Logistics rating	4.81	0.38	1	5	233,318
Natural experiment (Jun. 1–Jun. 3) 129,448 SKUs	Hour	Total sales	54,040.02	37,474.02	2,751	116,554	42
		Total "star" sales	39,149.62	27,969.41	1,760	87,421	42
		Product variety	8,958	4,586	1,490	14,264	42
	Product	Price	167.63	708.25	0.01	37,813.10	129,448
		Logistics rating	4.82	0.46	1	5	129,448

5.1. Difference-in-Differences Design

In our analysis, the starting time of the natural shock is 6:00 p.m. on June 1, and the ending time is 12:00 p.m. on June 3. This exogenous shock lasted for 42 hours, which we refer to as the treatment period.

The shock affected the delivery quality for some products but not all of them. Recall that only non-Cainiao merchants can choose SF as their delivery option, while Cainiao merchants can only use Cainiao logistics. That is, the exogenous shock affected only products sold through non-Cainiao merchants, which we define as our treatment group. Our data set includes transaction-level, merchant-level, and product-level information on Alibaba's retail platform before, during, and after the exogenous shock. This allows us to adopt a DID analysis to quantify the market outcomes—sales, product variety, sales dispersion, and logistics rating—driven by the treatment.

Our DID analysis compares treatment and control products across the treatment and non-treatment periods. A unique feature of our natural shock is that it involves both the decline and the reinstatement of delivery quality. We employ two DID approaches. First, following Parker et al. (2016), we combine pre-treatment, treatment, and post-treatment periods in one DID regression. The advantage of this approach is that the longer panel structure could help lower the standard error of estimates. Second, following Cui et al. (2018), we run a separate DID analysis to compare pre-treatment and treatment periods, and to compare treatment and post-treatment periods, respectively. The advantage of this approach is that we can obtain two sets of estimates: the loss from the removal of a high-quality logistics option (*pre-treatment comparison*) and the gain from the reinstatement of a high-quality logistics option (*post-treatment comparison*).

To make a relevant comparison, we analyze a 10-day window prior to the natural shock as our pre-treatment period—from May 20 to June 1 by 6:00 p.m.—and a seven-day window after the natural shock as our post-treatment period—from June 3 at 12:00 p.m. to June 10 by 11:59 p.m. In order to confirm that our results are not an artifact of this time specification, we replicate our analysis over two other time windows that span from May 15 to June 15, and from May 25 to June 8 as robustness checks in §6.1. Further, in order to avoid estimation biases, we discuss ways to eliminate potential confounding factors in §5.6.

5.2. Average Treatment Effect

Sales. To evaluate the impact of logistics on the platform, merchants, and products, we measure sales in three outcome variables: $Group\ Sales_{it}$, $Merchant\ Sales_{it}$, and $Product\ Sales_{it}$. $Group\ Sales_{it}$ is the total unit quantity sold in group i at hour t, where i represents the treated or control group. $Merchant\ Sales_{it}$ is the total sales quantity sold by merchant i at hour t, where i represents a particular merchant. $Product\ Sales_{it}$ is sales for product i at time t, where i represents a particular product, and t represents a time unit of six hours. Note that hourly sales at the product level has many zeros, and the size of the dataset causes a computation burden. To circumvent these issues, we average hourly sales into a six-hour window (note that the estimation results are almost identical if we aggregate the variable to a three-hour or two-hour window). In short, we exploit hourly variations in group and merchant sales, but six-hour variations in product sales.

In the first approach, to estimate the impact of high-quality logistics on the platform's sales, we compare the market outcomes of the treated products and the control products when SF is available on the platform against when SF is not available across pre-treatment, treatment, and post-treatment periods,

$$\begin{aligned} Sales_{it} = c + Treated_i + Treated\ Period_t + Post\ Period_t + \beta\ Treated_i \times Treated\ Period_t \\ + \gamma\ Treated_i \times Post\ Period_t + T_t + Z_i + e_{it}, \quad (1) \end{aligned}$$

where i denotes the group, merchant, or product; t denotes the number of hours or six-hour time units since the beginning of the observation period; T_t is the hour-of-the-day fixed effect; Z_i is a control vector that includes the average product price in i at time t and the product category fixed effect for product-level sales; e_{it} is the error term. The dummy variable $Treated_i$ equals 1 if product i belongs to the treatment group that could use SF prior to its removal and after its reinstatement. The dummy variable $Treated\ Period_t$ equals 1 if t belongs to the treatment period, and 0 otherwise. The dummy variable $Post\ Period_t$ equals 1 if t belongs to the post-treatment period, and 0 otherwise. The coefficient β estimates the value of logistics on sales, and the coefficient γ estimates the sales difference across the post-treatment and pre-treatment periods.

In the second approach, we run a DID regression across pre-treatment and treatment periods, or treatment and post-treatment periods,

$$Sales_{it} = c + Treated_i + Treated\ Period_t + \beta\ Treated_i \times Treated\ Period_t + T_t + Z_i + e_{it}, \quad (2)$$

where β estimates the treatment effect on sales.

Product variety and sales dispersion. We study two outcome variables: $Product\ Variety_{jt}$ and $Sales\ Dispersion_{jt}$. $Product\ Variety_{jt}$ is the number of unique SKUs sold in group j at hour t and $Sales\ Dispersion_{jt}$ is the contribution of long-tail sales to total sales in group j at hour t.

In the first approach, we compare product variety and sales dispersion when SF is available on the platform against when SF is not available across pre-treatment, treatment, and post-treatment periods,

$$\begin{aligned} Product\ Variety_{jt}\ (\text{or } Sales\ Dispersion_{jt}) &= c + Treated_j + Treated\ Period_t + Post\ Period_t \\ &+ \beta\ Treated_j \times Treated\ Period_t + \gamma\ Treated_j \times Post\ Period_t + T_t + Z_j + e_{jt}, \quad (3) \end{aligned}$$

where j denotes the treatment or control group; t denotes the number of hours; Z_j controls the average product price in group j at time t; T_t controls for the hour of the day. The coefficient β estimates the impact of low-quality logistics on product variety or sales dispersion, and the coefficient γ estimates the difference across post-treatment and pre-treatment periods.

In the second approach, we run a DID regression across pre-treatment and treatment periods, or treatment and post-treatment periods,

$$\begin{aligned} Product\ Variety_{jt}\ (\text{or } Sales\ Dispersion_{jt}) &= c + Treated_j + Treated\ Period_t \\ &+ \beta\ Treated_j \times Treated\ Period_t + T_t + Z_j + e_{jt}, \quad (4) \end{aligned}$$

where β estimates the treatment effect.

Logistics rating. Because not all orders receive logistics scores from consumers, the logistics rating at the hourly level has many missing values. To avoid the potential bias caused by the missing data, we aggregate the logistics scores for each product across three natural experiment periods, respectively. In particular, we average logistics ratings for product k in period p to obtain $Logistics\ Rating_{kp}$, where $p \in \{\text{pre-treatment period, treatment period, post-treatment period}\}$. Similarly, we first run a DID across pre-treatment, treatment, and post-treatment periods,

$$\begin{aligned} Logistics\ Rating_{kp} = c + Treated_k + Treated\ Period_p + Post\ Period_p + \beta\ Treated_k \times Treated\ Period_p \\ + \gamma\ Treated_k \times Post\ Period_p + Z_k + e_{kp}, \quad (5) \end{aligned}$$

where k denotes a particular product; $Treated\ Period_p$ is a dummy for being in the treatment period; $Post\ Period_p$ is a dummy for being in the post-treatment period; Z_k controls for the product category. The coefficient β estimates the impact of logistics quality on logistics rating, and the coefficient γ estimates the difference across post-treatment and pre-treatment periods.

The DID specification of the second approach is

$$Logistics\ Rating_{kp} = c + Treated_k + Treated\ Period_p + \beta\ Treated_k \times Treated\ Period_p + Z_k + e_{kp}, \quad (6)$$

where β estimates the treatment effect.

5.3. Effects Under Different Product Characteristics

Delivery quality might not affect all products equally, that is, the value of a high-quality delivery service is contingent on product-specific characteristics. We explore which products are more likely to be affected by logistics quality. In particular, we examine three moderating factors: popularity (i.e., star versus long-tail), product price, and price discount. To explore the moderating effects, we use a triple-difference specification to examine the differences in the impact of logistics on outcomes under different product attributes. We augment Equation (2) by including the moderating factors M_{it} and interacting them with the treatment effect,

$$\begin{aligned} Sales_{it} = c + Treated_i + Treated\ Period_t + Treated_i \times Treated\ Period_t \\ + \gamma\ M_{it} \times Treated_i \times Treated\ Period_t \\ + M_{it} \times Treated\ Period_t + M_{it} \times Treated_i + M_{it} + T_t + Z_i + e_{it}, \end{aligned} \tag{7}$$

where i denotes a particular product; t is the number of the time unit of six-hours; M_{it} is the moderator: when studying the popularity moderator, M_{it} is a dummy for being a popular item (M_{it} is reduced to M_i in this case, where $M_i = 1$ if product i is a star product and $M_i = 0$ otherwise); when studying the price moderator, M_{it} is the product price at hour t; when studying the discount moderator, M_{it} is the price markdown at hour t. The coefficient of interest is γ of the triple-interaction term. As before, we control for time, price, and product category fixed effects.

5.4. Summary Statistics for Treatment and Control Groups

Table 2 summarizes the statistics for the market outcomes—sales, variety, and logistics rating—and product attributes—price and discount—across the treatment and control groups in the pre-treatment, treatment, and post-treatment periods. From May 20 to June 10, 2017 on Alibaba's platform, 525 non-Cainiao merchants sold a total of 124,490 SKUs that could potentially be affected by the treatment, and 249 Cainiao merchants sold a total of 4,958 SKUs that were not affected by the treatment. Cainiao merchants hold fewer SKUs and their products are, on average, more expensive. We address this unbalanced nature by performing a DID with matching in §8.

Note that Alibaba, along with many retail companies, began running promotional events at 12:00 a.m. on June 1. This explains the spike in sales for the control group in the treatment period.[11] Table 2 provides initial evidence of the value of logistics: in the treatment period, the gap in sales between treated and control items significantly increased relative to both non-treatment periods, whereas the gap in product variety and logistics rating remained the same.

[11] If the value of high-quality logistics is significant, the treated products would have had the same significant increase in sales as the control products in the treatment period in the absence of the exogenous shock.

Table 2 Summary Statistics for Treatment and Control Groups

Time Window	Variable	Treatment Group (124,490 SKUs; 525 merchants) Mean	Std. dev	Control Group (4,958 SKUs; 249 merchants) Mean	Std. dev
Pre-treatment period (May 20–May 31)	Group sales	40,221.79	27,742.38	7,383.23	6053.74
	Merchant sales	77.55	546.48	31.99	151.35
	Product sales	0.32	21.13	1.61	18.20
	Product variety	7456	3736	982	472
	Logistics rating	4.82	0.58	4.86	0.47
	Price	135.96	628.78	1199.93	1926.17
	Discount (%)	−0.44	16.44	−0.25	10.97
Treatment period (Jun. 1–Jun. 3)	Group sales	41,087.40	29,785.20	12,952.62	8,021.65
	Merchant sales	78.94	556.71	52.02	125.52
	Product sales	0.33	21.02	2.61	17.88
	Product variety	7465	4001	1493	593
	Logistics rating	4.80	0.59	4.85	0.48
	Price	102.64	499.05	1068.69	1673.33
	Discount (%)	2.22	38.06	5.31	13.74
Post-treatment period (Jun. 4–Jun. 10)	Group sales	46,694.19	33,425.06	8,924.00	5,663.89
	Merchant sales	89.59	817.85	35.84	102.64
	Product sales	0.38	35.96	1.80	11.83
	Product variety	7810	3811	1220	498
	Logistics rating	4.81	0.58	4.85	0.48
	Price	118.78	565.43	1193.40	1976.31
	Discount (%)	−4.29	11.13	1.04	11.17

5.5. Parallel Trend Assumption

The critical identification assumption for the DID estimation is the parallel trends assumption: in the absence of the shock, the treatment and control groups would evolve in parallel. In order to justify the validity of this assumption, we test the time trends of pre-treatment sales at the product level. We need to eliminate the possibility of different pre-trends across the treatment and control groups. We examine the trend assumption by estimating the linear trend

$$Sales_{it} = c + \theta t + \lambda \ Treated_i \times t + T_t + Z_i + e_{it}, \tag{8}$$

where i denotes a particular product, and t is the number of six-hour time windows since the beginning of the observation period. We also control for time, price, and product category fixed effects. The coefficient λ captures the sales trend difference across treated and control products.

Table 8 presents the estimation result of Equation (8), and the estimated λ is statistically insignificant. Our results assure that there was no difference in pre-trends—this confirms that Cainiao merchants and non-Cainiao merchants evolved in parallel.

5.6. Potential Confounding Factors and Bias

In this section, we discuss and rule out several potential confounding factors that could bias our estimates.

Promotion in June. Alibaba, along with many retailers in China, runs promotions starting at 12:00 a.m. on June 1 every year. The promotions last until June 18. One might question whether the treatment group (non-Cainiao merchants) receives a different promotion scale than the control group (Cainio merchants) during the treatment period—Alibaba might want to promote more for merchants that choose to use its own logistics service.

We adopt three approaches to address this concern. First, in §6.1, we perform a DID analysis using price and price discount as outcome variables. We find that for both the pre-treatment comparison and the post-treatment comparison, the differences in price and discount across the treated and control products are statistically insignificant. Second, although Alibaba runs a promotion event in the treatment period, the level of promotion does not change much from the treatment period to the post-treatment period. In other words, the promotion intensity changed from May to June 1, whereas the promotion intensity did not change much from June 1 to June 10, which implies that the post-treatment analysis is unlikely to suffer from the confounding impact of promotions. Throughout the analyses, we find that all the estimates have almost the same magnitude and direction across the pre-treatment and post-treatment analyses, which further rules out the confounding impact of promotions. Last, we run a placebo test over another promotion period without the treatment. The DID estimates are insignificant, which implies that Alibaba promotes treated and untreated products equally.

Cannibalization. A potential concern of our study is that in the absence of a high-quality delivery option, customers might instead purchase from Cainiao merchants, which could cause a cannibalization issue. However, the cannibalization concern is less severe given that SF has a much stronger reputation among consumers than any other carriers served on Alibaba, including Cainiao logistics.[12] Consequently, even if customers are not satisfied with the low delivery quality, rather than buying from Cainiao merchants, they will likely switch to a competing retail platform with a service quality similar to SF, such as JD. Nonetheless, to better address this valid concern, in §8.2, we conduct a DID analysis with matching. In particular, we match each product with a "twin" product of similar price, discount level, and popularity but *from a different product category*. In short, the treatment affects only the treatment group and is unlikely to suffer from cannibalization.

Alibaba's potential strategic responses. Another potential concern is whether Alibaba strategically responded to the exogenous shock—SF's unexpected access block—by, for example, adjusting the ranking algorithm or pricing, promotion, and logistics strategies, which might confound the treatment. As a case in point, Alibaba might bump up the ranking of Cainiao merchants to prevent customers from noticing SF's disappearance from non-Cainiao merchants—thereby leading to the confounding of logistics' value with the attention generated by a new ranking algorithm. Alibaba might also allocate more resources to non-Cainiao merchants to minimize the damage to their sales or reputation.

[12] Founded in 2013, Cainiao logistics was still in the developing stage in 2017, not yet making the top 20 in the customer satisfaction survey (State Post Bureau 2017b).

However, the unexpected clash happened and ended so fast that companies had no time to respond. For example, the promotional events are usually organized in the following manner: Alibaba sets and announces promotional plans to merchants weeks or even months in advance; merchants apply for specific plans that fit them the best; in the meantime, merchants prepare their inventory and advertising accordingly. Changing the searching and ranking algorithms across Alibaba's entire platform not only takes a significant amount of time but also requires approvals from many levels—from vice presidents to the chief marketing officer. It also takes weeks or sometimes months to change logistics strategies. To further confirm these aspects, we conducted interviews with Alibaba's executives. To the best of their knowledge, Alibaba did not change pricing, promotion, searching, ranking, or logistics strategies during the clash, mostly due to the time constraint.

News. During the conflict between SF and Alibaba, the media propagated the story to potential customers, who would likely alter their shopping patterns accordingly and visit other platforms instead. However, news is unlikely to confound our treatment. First, the shock lasted only for 42 hours, which reduces the possibility of customers being influenced by news reports. Second, even if consumers decided to delay their purchase or visit another platform due to the news, this likely would have affected Cainiao and non-Cainiao merchants equally because the technical nuances between the two are not common knowledge among consumers. Last, we also estimate the treatment effect over the first 6, 12, 18, 24, and 36 hours of the clash and show that the effect sizes are highly consistent. In short, it is unlikely that media coverage confounds our estimations.

6. Estimation Results

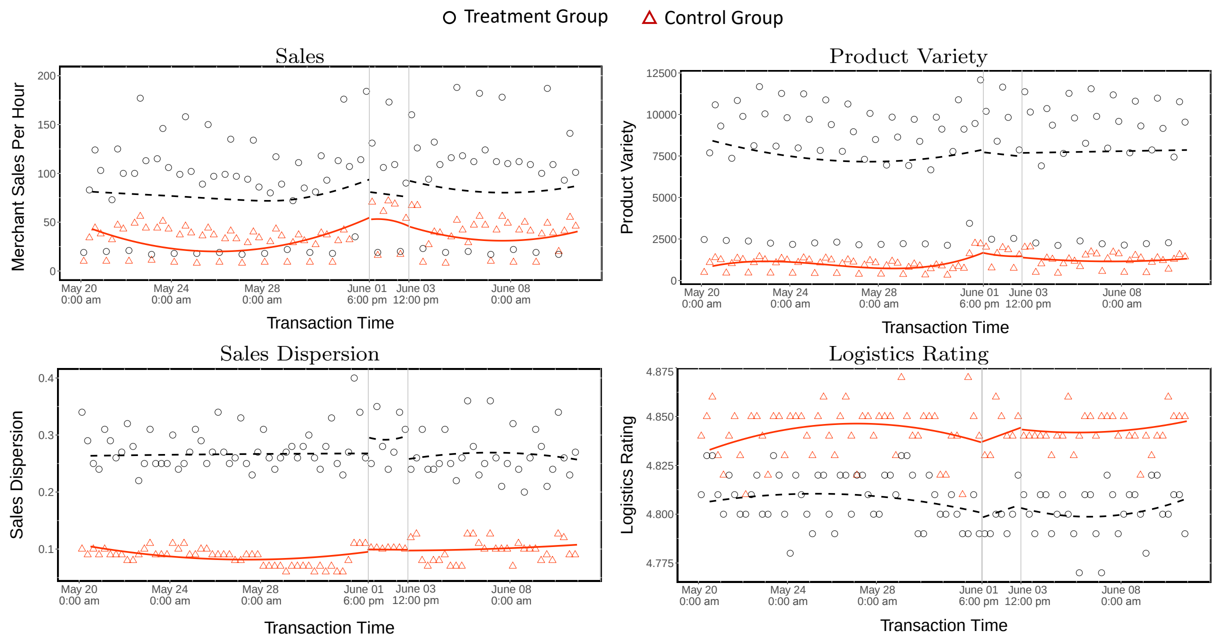
Our empirical analysis tests the hypotheses in §3. Here, we present the estimates for the outcome variables (sales, variety, sales dispersion, and logistics rating) and the moderators (popularity, price, and discount).

6.1. The Value of High-Quality Delivery

First, we analyze the overall effect of delivery quality on sales, product variety, sales dispersion, and logistics rating. In Figure 2, we plot the hourly trend between the treatment and control groups for each of the main outcome variables over the entire sample period. The two vertical lines indicate the starting and ending time of the shock. We can observe that the sales and sales dispersion of treated products are affected by the shock, whereas product variety and logistics ratings are not affected by the shock.

Table 3 depicts the estimates of the first DID approach in Equations (1), (3), and (5). We find a positive and statistically significant treatment effect on group sales (at 0.05 level) and merchant and product sales (at 0.001 level), which suggests that the removal of the high-quality delivery option

Figure 2 **Impact of Logistics Quality: Graphical Analysis**



Note: The dotted (solid) lines represent the fitted trend of the treatment (control) group using second order polynomials. Each dot is the outcome variable at the hourly level averaged within a six-hour window.

causally increases sales. In addition, we find insignificant coefficients of $Treated \times Post\ period$ for sales, which suggests that the resumption of SF resumes sales. Specifically, column I of Table 3 shows that hourly sales decrease by 7,501.14 units during SF's absence, which with a historical base of 51,503.58 hourly sales (in Table 1) represents a 14.56% drop in sales for the platform. Table 4 presents the treatment effects across pre- and post-comparisons following the second DID approach in Equations (2), (4), and (6), which are highly consistent with the above results. For example, column I of Table 4 presents an increase of 8,457.43 hourly sales, which represents a 16.42% drop in sales for the platform. Column II of Table 4 shows that the platform's sales resume with an increase of 9,695.71 hourly sales upon SF's resumption. These results confirm Hypothesis 1, thereby suggesting that consumers regard delivery quality as a key factor in their shopping decisions and react accordingly when a retail platform lacks a high-quality delivery option. Further, we explore the impact on the platform's and merchants' revenue, defined as product price times sales. The findings are presented in Table 9 in the appendix, showing a loss of ¥1.78 million in hourly revenue for the platform in the absence of SF.

Hypotheses 2 and 3 are also supported by columns IV–VI of Table 3 and columns VII–XII of Table 4. The reduction in delivery quality increased the contribution of long-tail sales to total sales by 3%, thereby flattening the sales dispersion. The estimates for product variety and logistics rating are not statistically significant. We run a robustness test where the outcome variable is the number of unique long-tail SKUs and, same as the number of unique SKUs, the estimates are insignificant. This suggests that the product variety is not affected by the delivery quality, because long-tail products—which determine the width of product lines—are less likely to be supplied and encroached by other platforms. Our results also show that the logistics rating is not affected. This indicates that customers who choose to continue to purchase during the shock are not particularly selective about delivery quality and, thus, do not perceive the quality difference.

Table 3 Value of Logistics Quality: DID across Pre-treatment, Treatment, and Post-treatment Periods

	Group Sales I	Merchant Sales II	Product Sales III	Product Variety IV	Sales Dispersion V	Logistics Rating VI
Treated	$23,348.34^{***}$	33.75^{***}	-2.25^{***}	$6,455.70^{***}$	0.27^{***}	-0.04^{***}
	(1864.55)	(1.31)	(0.25)	(141.50)	(0.005)	(0.007)
Treated period	$11,293.45^{***}$	27.81^{***}	0.29^{***}	765.30^{**}	-0.01^{*}	-0.009
	(2,222.40)	(2.87)	(0.08)	(287.00)	(0.006)	(0.01)
Post Period	2,319.68	3.42^{*}	0.04	131.80	0.0003	-0.004
	(1,245.33)	(1.70)	(0.09)	(163.80)	(0.003)	(0.009)
Treated × Treated period	$-7,501.14^{*}$	-20.42^{***}	-0.29^{***}	-398.40	0.03^{***}	0.002
	(3,079.54)	(3.54)	(0.08)	(405.60)	(0.008)	(0.01)
Treated × Post period	2,322.87	2.03	-0.08	105.50	0.008	-0.006
	(1,754.24)	(2.06)	(0.09)	(232.40)	(0.004)	(0.01)
Product controls	No	No	Yes	No	No	Yes
Time fixed effects	Yes	Yes	Yes	Yes	Yes	No
Observations	1,054	407,898	11,380,809	1,054	1,054	191,209
No. of SKUs	129,448	129,448	129,448	129,448	129,448	129,448
R^2	0.76	0.03	0.11	0.83	0.90	0.01

Note: This table reports the estimated coefficients and standard errors (in parentheses) in Equations (1),(3), and (5). The coefficients for group sales, merchant sales, product sales, product variety, sales dispersion, and logistics rating are presented in columns I–VI, respectively. Product controls include the price and the product category. Time fixed effects control the hours of the day. Significance at $^{*}p < 0.05$; $^{**}p < 0.01$; $^{***}p < 0.001$.

Time window robustness. To validate that our findings are not an artifact of the choice of time window, we examine two other time windows: May 15 to June 15 and May 25 to June 8. For both new time windows, we replicate our analysis over sales, sales dispersion, variety, and logistics rating for both DID analysis; the results reveal that the treatment effects retain the same sign and magnitude.

Robustness to promotion. To eliminate a potential confounding factor—unequal promotional intensity across Cainiao and non-Cainiao merchants prior to and after the shock—we test the dynamics of product price and discount. In particular, we run a DID analysis using price and discount as dependent variables, where we take their logarithm to account for their skewed distributions. The insignificant estimates presented in Table 10 in the appendix suggest that there is no systematic price or promotion difference across the treatment and control groups. This indicates that it is unlikely that Alibaba ran a promotion in June that particularly favored Cainiao or non-Cainiao merchants.

To further confirm this, we run a placebo test around the peak of Alibaba's 2017 Mid-Year Mega Sales—from June 15 to 18—when delivery quality did not change. In the absence of the treatment, if promotion is not a confounding factor, there should be no "treatment effect." This event has the same three-period structure as our studied clash: before the promotion peak (June 8 to June

Table 4 Value of Logistics Quality: DID for Pre-comparison and Post-comparison

	Group Sales		Merchant Sales		Product Sales	
	I. Pre	II. Post	III. Pre	IV. Post	V. Pre	VI. Post
Treated	$20,957.05^{***}$	$26,615.66^{***}$	33.68^{***}	35.54^{***}	-2.26^{***}	-2.33^{***}
	(5760.37)	(6739.99)	(1.29)	(1.73)	(0.01)	(0.02)
Treated period	$12,195.24^{***}$	$9,091.84^{***}$	27.72^{***}	24.95^{***}	0.29^{***}	0.26^{***}
	(1365.78)	(1173.22)	(2.82)	(3.07)	(0.03)	(0.03)
Treated × Treated period	$-8,457.43^{***}$	$-9,695.71^{***}$	-20.38^{***}	-22.56^{***}	-0.28^{***}	-0.21^{***}
	(1509.78)	(1585.61)	(3.48)	(3.78)	(0.03)	(0.03)
Product controls	No	No	No	No	Yes	Yes
Time fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Observations	694	444	268,578	171,828	7,501,215	4,784,833
No. of SKUs	129,448	129,448	129,448	129,448	129,448	129,448
R^2	0.76	0.76	0.03	0.03	0.83	0.84
	Product Variety		Sales Dispersion		Logistics Rating	
	VII. Pre	VIII. Post	IX. Pre	X. Post	XI. Pre	XII. Post
Treated	$6,455.40^{***}$	$6,562.00^{***}$	0.27^{***}	0.27^{***}	-0.05^{***}	-0.05^{***}
	(143.30)	(185.10)	(0.006)	(0.008)	(0.01)	(0.01)
Treated period	762.90^{**}	649.20^{*}	-0.01	-0.01^{**}	-0.001	-0.005
	(290.80)	(300.60)	(0.006)	(0.006)	(0.01)	(0.01)
Treated × Treated period	-398.10	-504.70	0.03^{***}	0.04^{***}	0.001	0.008
	(410.60)	(424.00)	(0.008)	(0.008)	(0.01)	(0.01)
Product controls	No	No	No	No	Yes	Yes
Time fixed effects	Yes	Yes	Yes	Yes	No	No
Observations	694	444	694	444	171,258	158,631
No. of SKUs	129,448	129,448	129,448	129,448	129,448	129,448
R^2	0.83	0.84	0.90	0.90	0.01	0.01

Note: This table reports the estimated coefficients and standard errors (in parentheses) for the pre-treatment analysis (Pre) and the post-treatment analysis (Post) in Equations (2),(4), and (6). The estimated coefficients for group sales, merchant sales, product sales, product variety, sales dispersion, and logistics rating are presented in columns I–II, III–IV, V–VI, VII–VIII, IX–X, and XI–XII, respectively. Product controls include the price and the product category. Time fixed effects control the hours of the day. Significance at $^{*}p < 0.05$; $^{**}p < 0.01$; $^{***}p < 0.001$.

14), during the promotion peak (June 15 to June 18), and after the promotion ends (June 19 to June 30). Table 11 in the appendix presents the DID analysis of the placebo test, which shows insignificant differences between Cainiao and non-Cainiao groups caused by promotions.

Robustness to news. To rule out the news effect, we estimate the treatment effect in sales using the first 6, 12, 18, 24, 30, 36, and 42 hours of the clash as the treatment period; the estimates are plotted in Figure 5 in the appendix. The news effect, if existed, would have diffused over time, but we do not observe a significant time trend and the estimates are consistent over time, thereby confirming that news is not a confounder.

6.2. Moderating factors

What strategic levers could the platform implement to strengthen the effect of delivery quality on sales? Next, we estimate the relative effect of delivery quality on sales as a function of product popularity, product price, and price markdown specified in Equation (7). Table 5 presents the moderating estimates for both pre-treatment and post-treatment comparisons. The estimates are, again, highly consistent in both direction and magnitude.

Popularity. Recall that a high-quality delivery boosts sales. Columns II and VI in Table 5 show that the product popularity amplifies this effect in both pre-treatment and post-treatment comparisons. To be specific, we find that star products sold 1.05 fewer units per hour compared to long-tail products when the delivery quality fell, and that star products sold 0.45 more units per hour when the high-quality delivery option was restored. We run a robustness test where star products are redefined to account for 50%, 60%, 70%, and 90% of total sales, and the results are robust across these variations. This is because popular products are also supplied by other retail platforms that customers can switch to; thus, they are more adversely affected by a decline in service quality.

Product price. Columns III and VII of Table 5 show that the product price amplifies the value of delivery quality in both pre-treatment and post-treatment comparisons. In particular, we find that a higher product price (i.e., one in the top quartile of the distribution) amplifies the treatment effect by 64.89% in sales as compared to a lower price (i.e., one in the bottom quartile of the distribution). This is because customers tend to seek a reliable option that will ensure the safe arrival of their expensive purchases. In the absence of a high-quality delivery option, customers who buy expensive products are more likely to turn to other retail platforms.

Discount. Columns IV and VIII of Table 5 show that the discount weakens the value of delivery quality in both pre-treatment and post-treatment comparisons. In particular, we find that a higher price markdown (i.e., one in the top quartile of the distribution) weakens the treatment effect by 46.08% in sales as compared to a lower price markdown (i.e., one in the bottom quartile of the distribution). This is because customers trade off the price markdown and the delivery quality when deciding whether to make a purchase. When benefiting from a deeper discount, customers are more likely to put up with mediocre delivery logistics. Thus, in this sense, price markdown and delivery quality are strategic substitutes.

7. Mechanism

One might ask whether the drop in sales during the clash is lost temporarily or permanently. Did customers postpone their purchase or buy somewhere else? In this section, we test consumer search behavior in §7.1 and consumer purchasing behavior in §7.2, the findings of which indicate that some customers were switching to buy from another retailer with a comparable logistics service as SF (such as JD). That is, Alibaba indeed lost consumers to competitors in the absence of SF.

7.1. Consumer Search

Search engines have been a primary tool for consumers to search for products, services, or information. A customer's proactive searches are considered as valuable indicators of her interests. Thus, consumer search volume has been shown to be a strong indicator of product demand (Liu and Toubia 2018). In our research context, when SF was not available on Alibaba, if customers were

Table 5 Value of Logistics and Moderators

	Hourly Sales at the SKU level							
	Pre-treatment Comparison				Post-treatment Comparison			
	I	II	III	IV	V	VI	VII	VIII
Treated	-2.26^{***}	-0.71^{***}	-3.53^{***}	2.48^{***}	-2.33^{***}	-0.74^{***}	-3.70^{***}	1.21^{***}
	(0.01)	(0.02)	(0.04)	(0.13)	(0.02)	(0.02)	(0.05)	(0.15)
Treated period	0.29^{***}	0.27^{***}	-0.09	4.22^{***}	0.26^{***}	0.23^{***}	-0.15	3.02^{***}
	(0.03)	(0.04)	(0.10)	(0.25)	(0.03)	(0.04)	(0.10)	(0.26)
Treated × Treated period	-0.28^{***}	-0.21^{***}	0.14	-4.18^{***}	-0.21^{***}	-0.14^{***}	0.16	-3.10^{***}
	(0.03)	(0.04)	(0.10)	(0.26)	(0.03)	(0.04)	(0.10)	(0.27)
Star product		3.89^{***}				4.46^{***}		
		(0.02)				(0.03)		
Star product × Treated × Treated period		-1.05^{***}				-0.45^{***}		
		(0.06)				(0.06)		
Price			-0.67^{***}				-0.70^{***}	
			(0.008)				(0.009)	
Price × Treated × Treated period			-0.08^{***}				-0.06^{**}	
			(0.02)				(0.02)	
Discount				6.62^{***}				5.24^{***}
				(0.18)				(0.21)
Discount × Treated × Treated period				5.61^{***}				4.29^{***}
				(0.36)				(0.38)
Product controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Time fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	7,501,215	7,501,215	7,501,215	7,501,215	4,784,833	4,784,833	4,784,833	4,784,833
No. of SKUs	129,448	129,448	129,448	129,448	129,448	129,448	129,448	129,448
R^2	0.11	0.23	0.08	0.10	0.11	0.23	0.08	0.11

Note: This table reports the estimated coefficients and standard errors (in parentheses) for moderators star product, product price, and discount, as specified in Equation (7). The dependent variable is hourly sales at the SKU level. Columns I–IV report the pre-treatment comparison analysis. Columns V–VIII report the post-treatment comparison analysis. Product controls include the price and the product category. Time fixed effects control the hours of the day. We have included single, double, and triple interactions in the specification. For simplicity, we only report the variables of interest. Significance at $^{*}p < 0.05$; $^{**}p < 0.01$; $^{***}p < 0.001$.

considering other buying options, they would express interests and search for competing retailers that offer good delivery service. Therefore, we expect that the search queries for Alibaba's key competitor's logistics service—i.e., JD logistics—would significantly increase during the clash, as compared to the search queries for carriers that were not affected by the clash.

We employ a DID analysis: the treated group is the search volume of JD logistics, and the control group is the search volume of STO Express (STO), which is one of the largest carriers in China with a middle-ranked delivery quality. To support this, we collect search data from a leading search engine that represents people's general search behavior. We collect hourly search data for the following Chinese keywords: "Jing Dong Kuai Di," "Jing Dong Wu Liu," "Jing Dong Song Huo," and "Jing Dong Pei Song" from 64,926,800 search query records spanning across the period from May 20, 2017 to June 10, 2017. The above four keywords translate to JD's logistics (or delivery) in Chinese, and we measure its search volume as the summation of the above keywords. We also collect the search volume for the keyword "Shen Tong" (STO). Table 6 summarizes the statistics: the search for JD logistics increased to 9,917 over the treatment period and dropped to 9,247 afterwards. This presents us with preliminary evidence that customers showed interest in JD's delivery and might consider buying from JD when SF is not available on Alibaba.

Table 6 Summary Statistics for Hourly Search Volumes

Time Window	Hourly Search	Mean	Std. dev	Min	Max	N
Pre-treatment period (May 20–May 31)	JD Logistics	7,165.73	3,873.60	605	14,633	306
	STO Express	117,210.42	71,010.28	6,298	251,859	306
Treatment period (Jun. 01–Jun. 03)	JD Logistics	9,917.43	5,936.32	1,061	18,268	42
	STO Express	100,112.17	63,792.72	7,639	196,974	42
Post-treatment period (Jun. 04–Jun. 10)	JD Logistics	9,247.68	4,877.22	737	17,084	180
	STO Express	114,343.56	64,763.98	7,727	207,131	180

Table 7 reports the DID estimates. The search volume for JD logistics significantly increased by 23.36% as opposed to STO during the shock.[13] This finding supports the consumer switching mechanism: if customers only cease or postpone purchases, it is unlikely that their search for JD logistics would change.

Table 7 Consumer Search Analysis

	Log Search Volumes		
	I. Combined	II. Pre	III. Post
Treated	-2.70^{***}	-2.70^{***}	-2.45^{***}
	(0.01)	(0.01)	(0.01)
Treated period	-0.07^{**}	-0.07^{**}	-0.03
	(0.02)	(0.02)	(0.02)
Post period	-0.03^{*}		
	(0.01)		
Treated × Treated period	0.46^{***}	0.45^{***}	0.21^{***}
	(0.03)	(0.04)	(0.03)
Treated × Post period	0.25^{***}		
	(0.01)		
Time fixed effects	Yes	Yes	Yes
Observations	1,054	694	444
R^2	0.99	0.99	0.99

Note: This table reports the estimated coefficients and standard errors (in parentheses). Column I reports the DID analysis with combined periods (Combined). Columns II–III report the DID analysis with the pre-treatment analysis (Pre) and the post-treatment analysis (Post). Time fixed effects control the hours of the day. Significance at $^{*}p < 0.05$; $^{**}p < 0.01$; $^{***}p < 0.001$.

7.2. Consumer Retention

When SF was removed from Alibaba, the existing customers—who had shopped and experienced SF's high-quality service before—were more likely to notice its disappearance and react to it. Therefore, we expect that the clash reduced the retention of existing customers. To test this, we measure the retention rate as the percentage of existing customers, who had purchased (the studied SKUs) before the beginning of our observation period. Then, we run the DID analysis at the group level using retention rate as the dependent variable; see Table 12 in the appendix. We find that

[13] The post-comparison is not confounded with promotions and, thus, its estimated treatment effect is more valid. Because the outcome variables are logged, the percentage change in JD logistics searching equals $e^{0.21} - 1 = 23.36\%$.

consumer retention decreased during the clash but did not resume thereafter. Recall that in §6.1, we find that the post-treatment sales did not exceed the pre-treatment sales. Again, these results support the consumer switching behavior: the customers lost during the clash are more likely to make their purchase elsewhere, and existing customers, disappointed by SF's absence, are less likely to revisit the platform.

8. Difference-in-Differences with Matching

A challenge in our empirical design is that products are not randomly assigned to receive the treatment. To clear this hurdle, we employ a DID analysis with matching in §8.1. This method further helps rule out the cannibalization effect—customers in the treatment group might switch to buy products from the control group during the shock—in §8.2.

8.1. Matching and Results

We apply the DID with matching design to obtain a balanced sample, which allows us to make an "apples-to-apples" comparison. To this end, we match a subset of thousands of products pairs that are equally expensive or popular but receive different delivery quality options. We then re-estimate the treatment effect using products with identical (similar) observable attributes.

We apply the Mahalanobis distance matching method (Ho et al. 2007). For each control product, we find a "twin" product with identical characteristics among all the treated products.[14] The twin product should have (i) the same product category, (ii) similar price, and (iii) similar pre-treatment sales. We use nearest-neighbor matching measured in the Mahalanobis distance. In particular, we implement one-to-one matching without replacement, pairing an observation with the closest neighbor. We use the caliper-matching estimator with a tolerance level of 0.1, which specifies the maximum distance between pairs. The matching process yields a subsample of 4,393 SKUs in each group. Table 13 in the appendix summarizes how the treatment and control groups are balanced across all observable attributes.

We perform a DID analysis using the matched sample. Table 14 in the appendix shows that the parallel trend assumption is satisfied. Table 15 in the appendix shows that the average treatment effects on sales across group, merchant, and product levels are significant and positive at 0.001, and the effect on logistics rating remains statistically insignificant. Further, the size and significance of the estimates are consistent to estimations without matching in §6.

[14] We match controls from the treated group because there are fewer products in the control group by Cainiao merchants.

8.2. Cannibalization

One potential identification issue that could lead to an overestimation is cannibalization—customers might opt to buy products from Cainiao merchants when SF is not available. We address this concern by re-running the analysis using a different sample that matches products from different categories but with similar price and pre-treatment sales. This new sample mimics a "cannibalization-free" situation, which is based on a reasonable assumption that customers aiming to buy one type of product will not instead buy another type of product just to use the Cainiao service. We obtain a subsample with 4,128 SKUs in each group. Table 16 reports the results: in the absence of cannibalization, the average treatment effects retain the same sign and magnitude.

Further, we obtain another subsample that requires no overlapping in categories across treatment and control groups. That is, we match control products with treated ones from categories that no control items belong to, which completely prevents any potential switching behavior across the treatment and control groups. This sample has 4,414 SKUs in each group, and Table 17 reports results that are consistent with Table 16. These results suggest that the potential cannibalization, even if it exists, has a minimal impact on our analysis.

9. Conclusions

It is important to understand whether the (intangible) benefits of improving logistics quality, if any, exceed the costs of doing so. By exploiting a clash between Alibaba and SF Express, this paper takes one of the first steps toward quantifying the economic value of logistics to sales, variety, sales dispersion, and customer satisfaction. We show that delivery quality has a significant impact on sales and sales dispersion but does not impact the product variety or the logistic rating. Moreover, the effect is contingent on product-specific characteristics: popularity, price, and price markdown. In particular, we find that product popularity attenuates the value of logistics, customers are more sensitive to logistics choices when purchasing expensive items, and that price markdown mitigates the loss in logistics quality.

To employ logistics as an effective competitive lever and as a significant component of strategy, management needs to take both tactical and strategic actions (Heskett 1977). It should not only support ongoing corporate strategies but also factor logistics into the integral design of operations strategies on a long-term basis, such as logistics strategy evaluation and logistics system design.

Further, our results provide tactical and strategic recommendations to retail platforms and merchants aimed at enhancing their value propositions: a higher delivery quality significantly drives consumers' decisions to buy. For such endeavors, we further render the following roadmap to execute at a strategic level. First, a retail platform should evaluate its current logistics performance as quantified by logistics rating. However, we suggest looking beyond this measure because the rating system may already filter out selective consumers. Further, our results can guide companies in (re-) designing their reputation system for logistics. A system that does not include lost consumers is biased. To obtain a logistics measure that better reflects logistics performance, a company should also collect opinions from non-purchasing consumers via questionnaires and surveys. The company can use such methods to better diagnose, investigate, and optimize its logistics system. Last, a retail platform should evaluate its product portfolios to see whether the high-quality logistics strategy fits. For example, companies offering fast-moving products, expensive products, or rarely marked down products should particularly prioritize high-quality logistics.

Our study has a few limitations. First, we study a treatment effect that eliminates a complete package of logistics services (e.g., delivery speed, delivery reliability, and customer service). As a whole, this bundle of delivery services proves to be important. However, our study is unable to pinpoint the effect of specific aspects of a delivery service. We hope that future research can investigate this issue further by exploiting the finer aspects of delivery quality (Fisher et al. 2018, Calvo et al. 2018, Cui and Sun 2018). Second, we study a natural shock that removes SF as a shipping option, which confirms the value of the best service provider. However, we are unable to extrapolate our results for other, such as second-best, quality levels. Future research can explore a full spectrum of service quality. Last, collecting new data on traffic and buyers' demographics would enable future research to investigate more nuanced heterogeneous treatment effects. For example, future work could study the value of logistics quality depending on different types of purchasing nature (e.g., impulse purchasing versus planned buying) or different types of consumers (e.g., income level, education, rural or urban area).

Acknowledgements: We acknowledge "2018 MSOM Data Driven Research Challenge" for granting us access to the data. We are grateful to the chief strategy officer of Cainiao, the chief strategy officer of SF Express, the chief executive officer of Hivebox, and the chief strategy officer of JD.com for their helpful comments. We thank Victor Martnez-de-Albniz, Tom Tan, Laura Wagner, the anonymous associate editor, and anonymous referees for their constructive and helpful feedback. We also appreciate the valuable feedback provided by colleagues who participated in seminars at the IESE Business School, Harvard Business School, Kellogg School of Management, University of Maryland, University of Notre Dame, and University of Minnesota.

References

Alibaba. 2017. Alibaba group announces march quarter 2017 and full fiscal year 2017 results. https://www.alibabagroup.com/en/news/press_pdf/p170518.pdf (accessed, March 15, 2019).

Alibaba. 2018. How to buy on Taobao.com. https://sea.taobao.com/howtobuy (accessed, March 15, 2019).

Bai, Yang. 2017. The clash between SF and Cainiao: Taobao sellers can not use SF (in Chinese). `http://tech.ifeng.com/a/20170602/44629414_0.shtml` (accessed, March 15, 2019).

Bloomberg News. 2017. Chinese billionaires clash over Alibaba's parcel deliveries. `https://www.bloomberg.com/news/articles/2017-06-02/chinese-billionaires-clash-over-alibaba-s-parcel-deliveries` (accessed, March 15, 2019).

Brynjolfsson, Erik, Yu Hu, Mohammad S Rahman. 2009. Battle of the retail channels: How product selection and geography drive cross-channel competition. *Management Science* **55**(11) 1755–1765.

Brynjolfsson, Erik, Yu Hu, Michael D Smith. 2010. Research commentarylong tails vs. superstars: The effect of information technology on product variety and sales concentration patterns. *Information Systems Research* **21**(4) 736–747.

Cachon, Gérard P, Santiago Gallino, Marcelo Olivares. 2018. Does adding inventory increase sales? Evidence of a scarcity effect in US automobile dealerships. Management Science, Forthcoming.

Calvo, Eduard, Victor Martínez-de Albéniz, A. Thiele. 2018. Fast or furious: The impact of delivery outcomes on future buyer behavior. Working paper, IESE Business School, Spain.

Chevalier, Judith A, Dina Mayzlin. 2006. The effect of word of mouth on sales: Online book reviews. *Journal of Marketing Research* **43**(3) 345–354.

Chinanews. 2017. Complete break-up: SF Express is removed from Alibaba. `http://www.chinanews.com/jingwei/06-01/44522.shtml` (accessed, March 15, 2019).

Cook, Thomas D, Donald Thomas Campbell, William Shadish. 2002. *Experimental and quasi-experimental designs for generalized causal inference*. Houghton Mifflin, Boston, MA.

Cottrell, Tom, Barrie R Nault. 2004. Product variety and firm survival in the microcomputer software industry. *Strategic Management Journal* **25**(10) 1005–1025.

Cui, Ruomeng, Tianshu Sun. 2018. Sooner or later? Learning from delivery speed information. Working paper, Emory University, GA.

Cui, Ruomeng, Dennis J Zhang, Achal Bassamboo. 2018. Learning from inventory availability information: Evidence from field experiments on Amazon. Management Science, Forthcoming.

Culey, Sean A. 2018. *Transition Point: From Steam to the Singularity*. Troubador Publishing, UK.

eMarketer. 2018. Alibaba, JD.com face increasing competition in China. `https://retail.emarketer.com/article/alibaba-jdcom-face-increasing-competition-china/5b44ed07ebd4000b24140989/` (accessed, March 15, 2019).

Fbicgroup. 2018. China retail e-commerce weekly updates. `https://www.fbicgroup.com/sites/default/files/WeeklyRetailUpdates_12-18Jan2018_En.pdf` (accessed, March 15, 2019).

Fisher, Marshall, Santiago Gallino, Joseph Xu. 2018. The value of rapid delivery in online retailing. Working paper, University of Pennsylvania, PA.

Gallino, Santiago, Antonio Moreno, Ioannis Stamatopoulos. 2016. Channel integration, sales dispersion, and inventory management. *Management Science* **63**(9) 2813–2831.

Heskett, James L. 1977. Logistics-essential to strategy. *Harvard Business Review* **55**(6) 85–96.

Hill, Terry. 2017. *Manufacturing strategy: The strategic management of the manufacturing function*. Macmillan, United Kingdom.

Ho, Daniel E, Kosuke Imai, Gary King, Elizabeth A Stuart. 2007. Matching as nonparametric preprocessing for reducing model dependence in parametric causal inference. *Political Analysis* **15**(3) 199–236.

Hong, Wang. 2017. *The Balanced Scorecard Study on the Corporate Social Responsibility of Electronic Commerce:—From the Case of Taobao Alibaba*. Emerald Publishing Limited, UK.

Joerss, Martin, Jrgen Schrder, Florian Neuhaus, Christoph Klink, Florian Mann. 2016. Parcel delivery: The future of last mile. `https://www.mckinsey.com/~/media/mckinsey/industries/travel%20transport%20and%20logistics/our%20insights/how%20customer%20demands%20are%20reshaping%20last%20mile%20delivery/parcel_delivery_the_future_of_last_mile.ashx` (accessed, March 15, 2019).

Kumar, Piyush, Manohar U Kalwani, Maqbool Dada. 1997. The impact of waiting time guarantees on customers' waiting experiences. *Marketing Science* **16**(4) 295–314.

Lambert, Douglas M, Renan Burduroglu. 2000. Measuring and selling the value of logistics. *The International Journal of Logistics Management* **11**(1) 1–18.

Leadingir. 2018. Analysis on the status quo and trend of express industry in 2018. `http://www.leadingir.com/hotspot/view/916.html` (accessed, March 15, 2019).

Lee, Cyrus. 2016. Alibaba achieves 3 trillion yuan transaction volume milestone. `https://www.zdnet.com/article/alibaba-achieves-3-trillion-yuan-transaction-volume-milestone` (accessed, March 15, 2019).

Liu, Jia, Olivier Toubia. 2018. A semantic approach for estimating consumer content preferences from online search queries. *Marketing Science* **37**(6) 930–952.

Liu, Ying, Tae-Won Kang. 2015. Operational model of china's express services industry: A case study of SF-Express. *Journal of International Trade & Commerce* **11**(5) 43–58.

Mcmillan, Alex Frew. 2018. Express delivery businesses are booming in China. `https://realmoney.thestreet.com/articles/12/15/2017/express-delivery-businesses-are-booming-china` (accessed, March 15, 2019).

Meola, Andrew. 2016. Chinese e-commerce companies are growing by leaps and bounds, but there's still one problem. `https://www.businessinsider.com/chinese-e-commerce-companies-are-growing-by-leaps-and-bounds-but-theres-still-one-problem-2016-5` (accessed, March 15, 2019).

Mottl, Judy. 2018. Why delivery is playing a starring role in the retail customer experience. `https://www.retailcustomerexperience.com/blogs/why-delivery-is-playing-a-starring-role-in-the-retail-customer-experience` (accessed, March 15, 2019).

Parker, Chris, Kamalini Ramdas, Nicos Savva. 2016. Is it enough? Evidence from a natural experiment in india's agriculture markets. *Management Science* **62**(9) 2481–2503.

Ramdas, Kamalini. 2003. Managing product variety: An integrative review and research directions. *Production and Operations Management* **12**(1) 79–101.

Ramdas, Kamalini, Marshall Fisher, Karl Ulrich. 2003. Managing variety for assembled products: Modeling component systems sharing. *Manufacturing & Service Operations Management* **5**(2) 142–156.

Ramdas, Kamalini, Taylor Randall. 2008. Does component sharing help or hurt reliability? An empirical study in the automotive industry. *Management Science* **54**(5) 922–938.

Randall, Taylor, Serguei Netessine, Nils Rudi. 2006. An empirical examination of the decision to invest in fulfillment capabilities: A study of internet retailers. *Management Science* **52**(4) 567–580.

Roth, Aleda V, William E Jackson III. 1995. Strategic determinants of service quality and performance: Evidence from the banking industry. *Management Science* **41**(11) 1720–1733.

Soo, Zen. 2017. Cainiao and SF Express to resume data sharing after China's State Post Bureau intervenes. `https://www.scmp.com/business/companies/article/2096792/cainiao-and-sf-express-resume-data-sharing-after-chinas-state` (accessed, March 15, 2019).

State Post Bureau. 2017a. The consensus between Cainiao and SF Express (in Chinese). `http://www.spb.gov.cn/xw/dtxx_15079/201707/t20170703_1205540.html` (accessed, March 15, 2019).

State Post Bureau. 2017b. Notice on the state post bureau courier service satisfaction survey results 2017 (in Chinese) `http://www.spb.gov.cn/xw/dtxx_15079/201801/t20180129_1479117.html` (accessed, March 15, 2019).

Statista. 2018a. Alibaba continues to lead retail e-commerce sales in China in 2018. `https://www.statista.com/chart/14717/alibaba-continues-to-lead-retail-e-commerce-sales-in-china-in-2018/` (accessed, March 15, 2019).

Statista. 2018b. Number of internet users in China from December 2006 to December 2017. `https://www.statista.com/statistics/265140/number-of-internet-users-in-china/` (accessed, March 15, 2019).

Talmassons, Matteo. 2016. Taobao marketplace. `https://hackernoon.com/taobao-marketplace-c7797581af6a` (accessed, March 15, 2019).

Tan, Tom Fangyun, Serguei Netessine, Lorin Hitt. 2017. Is tom cruise threatened? An empirical study of the impact of product variety on demand concentration. *Information Systems Research* **28**(3) 643–660.

Thomas, Lauren, Courtney Reagan. 2018. Watch out, retailers. this is just how big Amazon is becoming. https://www.cnbc.com/2018/07/12/amazon-to-take-almost-50-percent-of-us-e-commerce-market-by-years-end.html/ (accessed, March 15, 2019).

Tianchi. 2018. Cainiao MSOM data-driven research competition. https://tianchi.aliyun.com/competition/introduction.htm?spm=5176.100066.0.0.cab63d74jciN6&raceId=231623 (accessed, March 15, 2019).

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

Wagner, Laura, Eduard Calvo, Ruomeng Cui. 2019. Disclosing product availability in online retail. Working paper, Emory University.

Wang, Amanda, Philip Glamann. 2018. China's biggest courier firm could soon deliver parcels by drone. https://www.bloomberg.com/news/articles/2018-03-28/china-s-biggest-courier-firm-could-soon-deliver-parcels-by-drone (accessed, March 15, 2019).

Weise, Karen. 2018. Last-minute shoppers increasingly trust only Amazon to deliver. https://www.nytimes.com/2018/12/21/technology/amazon-last-minute-gifts-shopping.html (accessed, March 15, 2019).

Ye, Josh. 2017. Cainiao, SF Express in standoff over data, causing confusion among Chinese online shoppers. https://www.scmp.com/business/article/2096631/cainiao-sf-express-standoff-over-data-gumming-deliveries-chinas-online (accessed, March 15, 2019).

Zhu, Feng, Marco Iansiti. 2012. Entry into platform-based markets. *Strategic Management Journal* **33**(1) 88–106.

Zhu, Feng, David Lane. 2015. SF Express: From delivery to e-commerce. Harvard Business School Case 616-003.

Zhu, Feng, Qihong Liu. 2018. Competing with complementors: An empirical look at Amazon. com. Strategic Management Journal, Forthcoming.

Appendices

Figure 3 Ranking of Consumer Complaint Rate for China's Major Carriers

Company	Consumer Complaint Rates (in PPM)			
	Overall	Delays	Damaged/Lost	Delivery
UC Express	41.94	17.62	8.23	15.37
ZJS Express	17.92	5.25	4.20	8.14
EMS	16.88	7.18	3.94	5.33
STO Express	10.68	2.74	2.37	5.34
YTO Express	9.96	3.35	2.34	4.10
Ttkd Express	7.86	2.22	3.72	1.73
Best Express	5.53	1.79	1.92	1.72
ZTO Express	2.09	0.52	0.50	1.00
JD Logistics	1.16	0.34	0.35	0.41
SF Express	**0.90**	**0.23**	**0.31**	**0.25**
Average	6.96	2.43	1.81	2.51

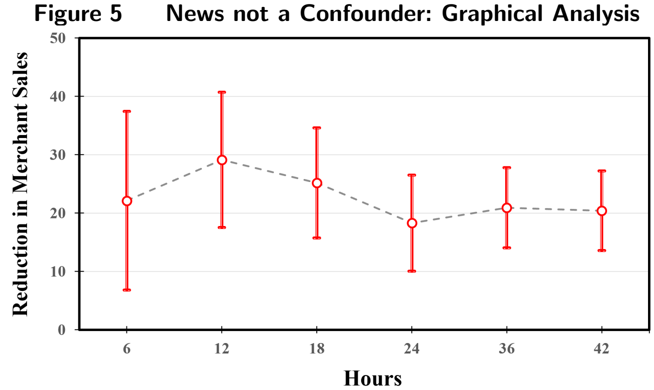
Note: PPM represents part per million. The table is translated from the survey available in http://www.askci.com/news/chanye/20180122/085409116522ˆ5.shtml.

Figure 4 Sales Dispersion Before and During the Clash

(a) Before the Clash (b) During the Clash

Note: The figures plot sales frequency over product popularity. The left (right) part of the plots represents long-tail (head) products. The clash increases sales dispersion.

Figure 5 News not a Confounder: Graphical Analysis



Note: The figure plots the treatment effect in sales over the first 6, 12, 18, 24, 36, and 42 hours of the clash. The vertical bands span 95% confidence intervals.

Table 8 Parallel Trend Test

	Hourly Sales at SKU Level
Treated	−1.28
	(0.75)
Treated × t	−0.03
	(0.02)
Product controls	Yes
Time fixed effects	Hourly
Observations	6,601,848
No. of SKUs	129,448
R^2	0.03

Note: This table reports the estimated coefficients and standard errors (in parentheses) in Equation (8). Product controls include the price and the product category. Significance at $^{*}p<0.05$; $^{**}p<0.01$; $^{***}p<0.001$.

Table 9 Value of Logistics Quality on Revenue (unit: ¥1,000)

	Group Revenue			Merchant Revenue		
	I.Combined	II.Pre	III.Post	IV. Combined	V.Pre	VI.Post
Treated	$6,490.52^{***}$	$7,466.21^{***}$	$5,842.89^{***}$	-1.04	-1.54^{***}	-0.92
	(331.02)	(436.47)	(456.62)	(1.30)	(0.27)	(0.62)
Treated period	$2,416.25^{***}$	$2,010.64^{***}$	$2,764.25^{***}$	10.09^{***}	12.31^{***}	7.82^{***}
	(394.54)	(433.26)	(342.46)	(1.43)	(0.59)	(1.10)
Post period	-99.30			5.38^{**}		
	(221.08)			(1.75)		
Treated × Treated period	$-2,177.68^{***}$	$-1,776.51^{**}$	$-2,533.81^{***}$	-9.42^{***}	-11.73^{***}	-8.17^{***}
	(546.71)	(591.50)	(475.52)	(1.30)	(0.73)	(1.36)
Treated × Post period	197.51			-3.54		
	(311.43)			(2.34)		
Time fixed effects	Hourly	Hourly	Hourly	Hourly	Hourly	Hourly
Observations	1,054	694	444	407,898	268,578	171,828
No. of SKUs	129,448	129,448	129,448	129,448	129,448	129,448
R^2	0.72	0.71	0.83	0.02	0.03	0.02

Note: This table reports the estimated coefficients and standard errors (in parentheses) using the DID approach with combined periods (Combined) and the DID approach with the pre-treatment analysis (Pre) and the post-treatment analysis (Post). The estimates for group revenue and merchant revenue are presented in columns I–III and columns IV–VI, respectively. Significance at $^*p < 0.05$; $^{**}p < 0.01$; $^{***}p < 0.001$.

Table 10 No Difference in Price and Price Discount across Treatment and Control Groups

	Price			Price Discount		
	I. Combined	II. Pre	III. Post	IV. Combined	V. Pre	VI. Post
Treated	-0.54^{***}	-0.53^{***}	-0.60^{***}	0.0003	0.0002	-0.00008
	(0.02)	(0.02)	(0.01)	(0.004)	(0.0004)	(0.0006)
Treated period	-0.08^{***}	-0.08^{***}	-0.08^{***}	0.001^{**}	0.002^{*}	0.001
	(0.02)	(0.02)	(0.02)	(0.0005)	(0.0006)	(0.0007)
Post period	-0.002			0.0004		
	(0.02)			(0.0005)		
Treated × Treated period	0.02	0.02	0.04	-0.001	-0.001	-0.0007
	(0.02)	(0.02)	(0.02)	(0.001)	(0.0006)	(0.0007)
Treated × Post period	-0.02			-0.001		
	(0.02)			(0.001)		
Product Controls	Yes	Yes	Yes	Yes	Yes	Yes
Observations	270,005	171,202	158,588	270,005	171,202	158,588
No. of SKU	129,448	129,448	129,448	129,448	129,448	129,448
R^2	0.65	0.66	0.66	0.003	0.003	0.003

Note: This table reports the estimated coefficients and standard errors (in parentheses) with price and price discount as dependent variables using the DID approach with combined periods (Combined) and the DID approach with the pre-treatment analysis (Pre) and the post-treatment analysis (Post). The estimates for price and price discount are presented in columns I–III and columns IV–VI, respectively. Significance at $^*p < 0.05$; $^{**}p < 0.01$; $^{***}p < 0.001$.

Table 11 Promotion not a Confounding Factor: Placebo Test

		Group Sales	
	I. Combined	II. Pre	III. Post
Treated	$56,703.07^{***}$	$58,399.06^{***}$	$72,239.12^{***}$
	(4,100.84)	(5,543.03)	(4,280.04)
Treated period	$16,058.44^{***}$	$16,453.59^{***}$	$9,646.13^{**}$
	(3,599.93)	(4,594.60)	(3,615.95)
Post period	5,383.65		
	(2,793.12)		
Treated × Treated period	1,885.81	1,467.41	4,946.36
	(5,054.88)	(6,438.72)	(5,110.82)
Treated × Post period	$-1,898.50$		
	(3,886.66)		
Time fixed effects	Yes	Yes	Yes
Observations	1,104	528	768
No. of SKU	135,190	135,190	135,190
R^2	0.51	0.42	0.50

Note: This table reports the estimated coefficients and standard errors (in parentheses). Column I reports the placebo test by the DID approach with combined periods (Combined). Columns II–III use the DID approach with the pre-treatment analysis (Pre) and the post-treatment analysis (Post). Significance at $^*p < 0.05$; $^{**}p < 0.01$; $^{***}p < 0.001$.

Table 12 Consumer Retention Analysis

	Consumer Retention Rate		
	I.Combined	II.Pre	III.Post
Treated	− 0.003**	− 0.003***	− 0.009***
	(0.001)	(0.001)	(0.001)
Treated period	−0.002	−0.002	0.007***
	(0.002)	(0.002)	(0.001)
Post period	− 0.009***		
	(0.001)		
Treated × Treated period	−0.008**	−0.008*	−0.002
	(0.003)	(0.003)	(0.003)
Treated × Post period	− 0.006***		
	(0.002)		
Time fixed effects	Hourly	Hourly	Hourly
Observations	1,056	696	444
No. of SKUs	129,448	129,448	129,448
R^2	0.51	0.41	0.56

Note: This table reports the estimated coefficients and standard errors (in parentheses) by the DID approach with combined periods (Combined) and the DID approach with the pre-treatment analysis (Pre) and post-treatment analysis (Post). Significance at $^*p < 0.05$; $^{**}p < 0.01$; $^{***}p < 0.001$.

Table 13 Summary Statistics after Matching

		Treatment Group 4,393 SKUs; 458 merchants		Control Group 4,393 SKUs; 245 merchants	
Time window	Variable	Mean	Std. dev	Mean	Std. dev
Pre-treatment period (May 20–May 31)	Group sales	6,074.51	3,807.84	6,277.13	5,361.21
	Merchant sales	13.29	69.88	27.88	135.55
	Product sales	1.39	11.88	1.56	15.69
	Logistics rating	4.83	0.38	4.86	0.24
	Price	995.10	1,427.01	1,077.69	1,463.01
	Discount (%)	−0.50	15.74	0.78	10.88
Treatment period (Jun. 1–Jun. 3)	Group sales	5,434.71	3,545.65	11,376.69	7128.70
	Merchant sales	11.86	55.19	46.44	106.65
	Product sales	1.23	9.63	2.60	14.80
	Logistics rating	4.82	0.41	4.85	0.30
	Price	648.04	1,166.42	1,015.98	1,425.75
	Discount (%)	3.40	16.94	5.29	13.89
Post-treatment period (Jun. 4–Jun. 10)	Group sales	5,478.99	3,343.21	7,950.38	5,122.28
	Merchant sales	11.96	72.71	32.45	94.30
	Product sales	1.24	12.52	1.81	10.52
	Logistics rating	4.83	0.33	4.85	0.26
	Price	787.55	1327.91	1068.29	1450.40
	Discount (%)	0.05	13.48	1.37	10.71

Table 14 Parallel Trend Test Using Matched Sample

	Hourly Sales at SKU Level
Treated	−0.05
	(0.25)
Treated × t	−0.003
	(0.002)
Product controls	Yes
Time fixed effects	Hourly
Observations	2,688,516
No. of SKUs	8,786
R^2	0.03

Note: Using the matched sample, this table reports the estimated coefficients and standard errors (in parentheses) in Equation (8). Product controls include the price and the product category. Significance at $^*p < 0.05$; $^{**}p < 0.01$; $^{***}p < 0.001$.

Table 15 Value of Logistics Quality Using Matched Sample

	Group Sales		Merchant Sales		Product Sales		Logistics Rating	
	I. Pre	II. Post	III. Pre	IV. Post	V. Pre	VI. Post	VII. Pre	VIII. Post
Treated	$4,437.75^{***}$	$1,714.84^{***}$	-12.36^{***}	-20.49^{***}	-0.088^{***}	-0.47^{**}	-0.03^{***}	-0.03^{***}
	(302.15)	(348.34)	(0.34)	(0.46)	(0.001)	(0.02)	(0.01)	(0.01)
Treated period	$2,247.22^{***}$	$2,979.03^{***}$	22.16^{***}	15.98^{***}	5.82^{***}	5.19^{***}	−0.01	−0.006
	(419.61)	(318.78)	(0.79)	(0.87)	(0.05)	(0.06)	(0.01)	(0.002)
Treated ×	$-2,141.58^{***}$	$-2,649.61^{***}$	-22.21^{***}	-14.08^{***}	-0.86^{***}	-0.55^{***}	0.002	−0.006
Treated period	(578.07)	(438.53)	(0.98)	(1.08)	(0.08)	(0.10)	(0.01)	(0.01)
Product controls	No	No	No	No	No	No	Yes	Yes
Time fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	No	No
Observations	694	444	243,941	156,066	3,048,742	1,950,492	13,568	11,955
No. of SKUs	8,786	8,786	8,786	8,786	8,786	8,786	8,786	8,786
R^2	0.78	0.88	0.03	0.04	0.02	0.02	0.03	0.03

Note: Using the matched sample, this table reports the estimated coefficients and standard errors (in parentheses) for the pre-treatment analysis (Pre) and the post-treatment analysis (Post). The estimates for group sales, merchant sales, product sales, and logistics rating are presented in columns I–II, III–IV, V–VI, and VII–VII, respectively. Significance at $^{*}p < 0.05$; $^{**}p < 0.01$; $^{***}p < 0.001$.

Table 16 In the Absence of Cannibalization: Robust to Cross-Category Matching

	Group Sales		Merchant Sales		Product Sales		Logistics Rating	
	I. Pre	II. Post	III. Pre	IV. Post	V. Pre	VI. Post	VII. Pre	VIII. Post
Treated	$4,375.48^{***}$	$1,748.68^{***}$	-11.84^{***}	-19.42^{***}	-0.29^{***}	-0.90^{**}	-0.03^{***}	-0.04^{***}
	(308.33)	(354.59)	(0.35)	(0.52)	(0.02)	(0.04)	(0.01)	(0.01)
Treated period	$2,463.73^{***}$	$2,871.12^{***}$	21.53^{***}	15.56^{***}	5.74^{***}	5.05^{***}	−0.01	−0.01
	(423.19)	(334.85)	(0.82)	(0.96)	(0.06)	(0.09)	(0.01)	(0.01)
Treated ×	$-1,923.61^{***}$	$-2,211.96^{***}$	-20.59^{***}	-13.02^{***}	-0.75^{***}	-0.41^{**}	0.004	0.004
Treated period	(580.61)	(460.76)	(1.01)	(1.19)	(0.11)	(0.15)	(0.01)	(0.01)
Product controls	No	No	No	No	No	No	Yes	Yes
Time fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	No	No
Observations	694	444	244,288	156,288	2,864,832	1,832,832	12,715	11,302
No. of SKUs	8,256	8,256	8,256	8,256	8,256	8,256	8,256	8,256
R^2	0.76	0.86	0.03	0.03	0.02	0.02	0.04	0.04

Note: Using the matched sample, this table reports the estimated coefficients and standard errors (in parentheses) for the pre-treatment analysis (Pre) and the post-treatment analysis (Post). The estimates for group sales, merchant sales, product sales, and logistics rating are presented in columns I–II, III–IV, V–VI, and VII–VII, respectively. Significance at $^{*}p < 0.05$; $^{**}p < 0.01$; $^{***}p < 0.001$.

Table 17 In the Absence of Cannibalization: Robust to Nonoverlapping Cross-Category Matching

	Group Sales		Merchant Sales		Product Sales		Logistics Rating	
	I. Pre	II. Post	III. Pre	IV. Post	V. Pre	VI. Post	VII. Pre	VIII. Post
Treated	-688.30^{*}	−196.00	16.36^{***}	33.14^{***}	-1.95^{***}	-2.29^{**}	-0.05^{***}	-0.05^{***}
	(326.30)	(692.80)	(2.42)	(5.66)	(0.05)	(0.23)	(0.01)	(0.01)
Treated period	$6,150.00^{***}$	$5,179.70^{***}$	25.48^{***}	23.80^{***}	1.86^{***}	1.46^{***}	−0.01	−0.01
	(664.50)	(1,130.20)	(3.64)	(6.91)	(0.09)	(0.28)	(0.01)	(0.01)
Treated ×	$-5,254.80^{***}$	$-5,747.10^{***}$	-20.28^{**}	-37.10^{**}	-0.99^{***}	-1.29^{**}	0.01	0.01
Treated period	(938.00)	(1,592.90)	(6.49)	(12.32)	(0.13)	(0.43)	(0.01)	(0.01)
Product controls	No	No	No	No	No	No	Yes	Yes
Time fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	No	No
Observations	694	444	171,765	109,890	3,063,316	1,959,816	120,922	109,185
No. of SKUs	8,828	8,828	8,828	8,828	8,828	8,828	8,828	8,828
R^2	0.60	0.47	0.01	0.004	0.01	0.001	0.01	0.01

Note: Using the matched sample, this table reports the estimated coefficients and standard errors (in parentheses) for the pre-treatment analysis (Pre) and the post-treatment analysis (Post). The estimates for group sales, merchant sales, product sales, and logistics rating are presented in columns I–II, III–IV, V–VI, and VII–VII, respectively. Significance at $^{*}p < 0.05$; $^{**}p < 0.01$; $^{***}p < 0.001$.