Assessing the Environmental Impact of Fast Shipping: Towards Sustainable Ranking Solutions*

Malika Korganbekova
The University of Chicago, malika.korganbekova@chicagobooth.edu

Aliya Korganbekova
Harvard Business School, akorganbekova@hbs.edu

Alibek Korganbekov
Boston University, alibekk@bu.edu

Vinny DeGenova
Wayfair LLC

Yasaman Khazaeni
Wayfair LLC

Cole Zuber
Wayfair LLC

Policymakers have expressed concerns about the impact of widespread fast-delivery promises in e-commerce, leading to inefficient shipping practices such as partially filled shipments. This may result in additional trips and higher transportation- and packaging-related emissions. In this paper, first, we estimate the causal effect of fast delivery promises on consumer outcomes and environmental emissions. Large-scale experiment conduced at online retailer Wayfair revealed unintended consequences when fast-delivery promises were removed. Specifically, we show that consumers were unable to easily access fast-delivered products located closer to them, resulting in a shift towards purchasing products located further away. This led to a 1% increase in environmental emissions, 2.4% longer delivery times, and 2.8% higher shipping costs. Next, considering the ineffectiveness of removing fast-delivery promises, we propose and evaluate a new product ranking policy to reduce emissions. Our proposed multi-armed bandit algorithm balances two goals: (i) it minimizes the shipping distance to each consumer by placing products located geographically closer to a consumer higher up in the ranking page results, and (ii) it maximizes the conversion probability. We show that the algorithm reduces emissions by 7.84% without compromising conversion or profit metrics.

Key words: environmental emissions, fast shipping, online retail, consumer nudges, multi-armed bandit

---

* We are grateful to Eric Anderson, Michael Dinerstein, Brett Gordon, Robert Porter, and Mar Reguant for helpful discussions. We also thank participants of the Conference on Digital Experimentation at Massachussetts Institute of Technology, Rising Scholars Conference at UChicago, AMA Sheth Doctoral Consortium at UT Austin, Trans-Atlantic Doctoral Conference at London Business School for helpful comments and suggestions. All remaining errors are our own. The authors gratefully acknowledge Wayfair for generously sharing the data for this project. Yasaman Khazaeni

and Cole Zuber were full-time employees at Wayfair during the research project, but are unaffiliated for research purposes.

1. Introduction

Online retail has grown significantly in recent years, in part due to the convenience of shopping from anywhere and having goods delivered directly to one's doorstep. Many consumers have come to expect fast shipping as a standard part of the online shopping experience, with 80% of consumers now expecting delivery within two days.[1] Policymakers are concerned that fast delivery promises negatively impact the environment through inefficient shipping practices: especially that the platform might send shipments that are only partially full to keep delivery promises.[2] This may result in additional trips and higher transportation- and packaging-related emissions. Despite the policy concerns, there is limited empirical evidence of the impact of fast shipping promises on environmental emissions, such as those related to greenhouse gases (GHG).

This paper assesses the causal effects of fast delivery promises on consumer choices and shipping outcomes, e.g., miles traveled by packages and resulting environmental emissions. To this end, we collaborated with Wayfair, a large US-based online retailer of home goods and furniture. To quantify the causal effects of fast delivery promises on consumer choices, we use a large-scale field experiment in which Wayfair randomly removed fast delivery promises from the product ranking pages consumers saw. The experiment ran for 1.5 months and spanned the entire website traffic. Consumers in the control group saw business-as-usual product rankings with all existing fast delivery promises. For the treated consumers, the platform removed 25-75% of fast delivery promises (also called *fast shipping badges*), pushed back products' get-it-by dates and changed the rankings of the affected products.

The large-scale field experiment indicates that consumers who did not see fast delivery promises were 7% less likely to choose fast-shipped products located closer to them, and instead substituted to products that were located in warehouses farther from them. This led to a 1% increase in the miles the product ultimately traveled, 2.4% longer delivery times, and 2.8% higher shipping costs in the treatment group compared to the control (business-as-usual) group. Using FedEx and UPS tracking data, we further document that given treatment group consumers' orders, the platform had limited ability to consolidate items for shipping, as measured by the number of orders that leave a warehouse at a given point in time. We find that this inability to consolidate orders led to an additional 2.5% increase in environmental emissions. Thus, removing fast shipping badges had a negative effect on consumer's delivery experience, shipping costs, and environmental emissions.

Having established that removing fast delivery promises does not decrease environmental emissions, we propose an alternative multi-armed bandit-based product ranking policy to help

[1] See 2020 Flexe Omnichannel Consumer Survey and McKinsey Report.

[2] See the CNN article, for example.

online retail platforms reduce their environmental emissions. The proposed ranking policy balances two goals: (i) it *minimizes the shipping distance* to each consumer by placing products located geographically closer to a consumer higher up in the ranking page results, and (ii) it *maximizes conversion probability*. We build on the idea of a linear UCB (Upper Confidence Bound) algorithm that selects which products (arms) to show to the consumers given the expected utility (reward) of each product and the ability to explore other products. The algorithm calculates an index for each product given its past performance when shown to the consumers, such as the click-through rate. In each iteration, the algorithm updates its posterior index of the products and shows consumers the products with the highest indices.

To evaluate the proposed ranking policy and quantify consumer welfare, we estimate a sequential search model (Weitzman 1979, Ursu 2018) that captures consumer search behavior on the website and preferences towards observable product characteristics, such as price, ratings, and fast delivery promises. We estimate the model on the data from the experiment described above. The experiment introduces exogenous variation in the (i) visibility of the fast shipping badges, (ii) product rankings and (iii) delivery day estimates. This variation allows us to identify the model parameters.

In the counterfactual simulations, we modify the product rankings that consumers see on the website according to the proposed ranking algorithm. We show the same rankings to all consumers living in a given state. In the simulations, consumers arrive one by one and sequentially search for a product. Information about previous consumers' clicks and purchases is propagated to the algorithm so that it can update posterior indices of products' relevance to the consumer. Thus, if the algorithm started by recommending an item that is eventually not clicked or purchased, the algorithm downgrades that product's index and stops showing it to subsequent consumers. Consumers either leave the website or purchase at the end of the search process.

A unique aspect of our data that makes counterfactual simulations realistic is that we observe inventory information and a history of supplier fulfillment rejection logs on the platform. The latter show both which supplier was selected to fulfill each historical order for a particular Zip Code and the full list of rejected suppliers. This information enables us to credibly identify which supplier is most likely to fulfill the order. In the cases where the number of orders of a product exceeded average inventory levels of the favored supplier, we chose the historically next best supplier.

Using the simulations, we show that the proposed ranking algorithm achieves lower environmental emissions and reduces shipping costs by 7.84% without compromising conversion rates. However, we also document that there is significant heterogeneity among states, with some states achieving significant savings while others do not.

To the best of our knowledge, this paper is among the first to estimate the causal impact of fast delivery promises on consumers and environmental emissions. Identifying the causal effects of fast

delivery can be challenging, but we are able to approach this challenge effectively using a large-scale field experiment conducted at Wayfair. Having established that removing fast delivery promises increases environmental emissions rather than reducing them, we propose an algorithm that can nudge consumers towards products located closer to them without diminishing their experience on the platform.

This paper contributes to the growing literature on climate change and sustainability. A paper closely related to ours is Muñoz-Villamizar, Velázquez-Martínez, Haro, Ferrer and Mariño (2021), where the authors run simulations to understand how fast shipping affects the product routing from an operations research perspective. They find that fast shipping reduces a platform's ability to consolidate the items in the mattresses category. Our experimental results suggest the opposite, and we think that it is crucial to account for consumer response in the absence of fast shipping promises, which Muñoz-Villamizar, Velázquez-Martínez, Haro, Ferrer and Mariño (2021) do not.

In the marketing literature, few papers assess the nudging of consumers towards sustainable options. For instance, Brecko and Kim (2023) and Bollinger, Kronthal-Sacco and Zhu (2023) study the provision of sustainability claims in the packaged goods market. Generally, it is crucial to deliver substantiated claims and raise consumer awareness about sustainability information. In our counterfactuals, we take a different route: we nudge consumers towards more environmentally friendly products by re-ranking products such that these products are more prominent. However, in the counterfactuals consumers do not know the logic behind the rankings.[3]

In operations research and transportation fields, there is a large literature that focuses on optimizing product routing and consolidation to reduce emissions. For example, Kang, Song, Xu, Miller, Wang, Zhang, Liu, Zhou, Ren, Zhong et al. (2021) propose low-carbon pathways for the express delivery in China; Tran and Lam (2022) and Cristea, Hummels, Puzzello and Avetisyan (2013) analyze container-ship trade flows; Juan, Faulin, Pérez-Bernabeu and Jozefowiez (2014) show that there is a potential to decrease greenhouse emissions if multiple firms along the supply chain collaborate to fulfill the orders jointly; and Van Loon, Deketele, Dewaele, McKinnon and Rutherford (2015) perform a comparative analysis of carbon emissions in the retail setting.

Finally, there is a growing body of research in economics on the environmental regulation of oil, gas, and electricity markets (Kellogg and Reguant 2021) and transportation (Cristea, Hummels, Puzzello and Avetisyan 2013). A tangentially related literature concerns reducing grocery-store waste (Sanders 2020) and reducing post-consumer packaging waste (Duan, Song, Qu, Dong and Xu 2019).

[3] It would be interesting to run field experiment with the platform where consumers are notified about the logic behind product rankings, but we leave it for future research.

Overall, to the best of our knowledge, ours is the first empirical paper that uses experimental results to investigate the environmental impact of fast shipping. Our proposed ranking algorithm also relates to the vast literature on contextual multi-armed bandit algorithms (Auer, Cesa-Bianchi and Fischer 2002 and Li, Chu, Langford and Schapire 2010).

The rest of the paper is organized as follows. Section 2 describes the empirical setting and the data. Section 3 describes the fast shipping experiment and presents our main experimental results. Section 4 proposes alternative ranking algorithms. Section 5 presents the structural model of consumer search and explains identification and estimation results. We present counterfactual results in Section 6. Section 7 concludes.

2. Empirical Setting

In this paper, we use data generously provided by Wayfair (hereafter, the "platform"), a large US-based online retailer of furniture and home goods. Wayfair does not own sellers' inventory, instead acting as an intermediary between product sellers and consumers. In this section, we walk the reader through the platform's interface design, explain typical consumer search process and the role of fast delivery promises.

Figure 1 Product Fast Shipping Badge

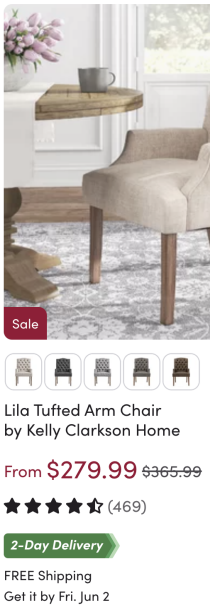
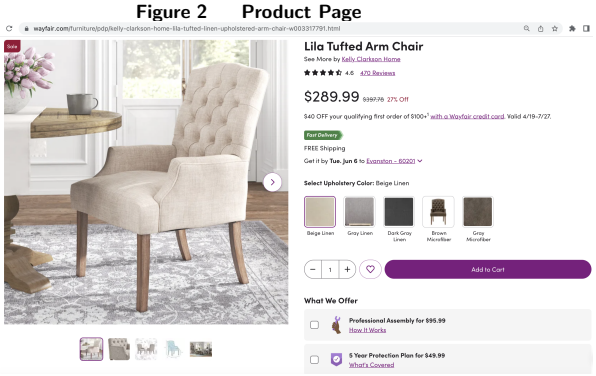


Figure 2 Product Page



When consumers arrive to the platform, they can use keyword box to input a search query or can indicate which product category they are interested in by choosing one of the existing product categories, e.g. chairs, kitchen appliances. Suppose consumers search for dining chairs. They are directed to the product ranking pages where each ranking page typically features 48 products.

Products are arranged in a matrix of 12×4. On the ranking pages, platform shows product images, sale price, ratings, numbers of ratings, and fast shipping badges (if applicable). Fast shipping badges are the green tags below the product's price that indicate the number of days in transit for delivery, such as 1-day or 2-day (see Figure 1). These products typically have free shipping. Below the badge, consumers also see the expected calendar delivery day.

Note that the badges are only shown to the consumers that are located within the 2-day delivery radius of the product.[4] Thus, same product in Figure 1 will not have the badge for the consumers whose network-implied geographic location is outside the seller's 2-day delivery area, e.g. Hawaii. The platform automatically adjusts the badge's visibility to ensure that sellers can fulfill the item delivery within one or two days.

If consumers like a product on the ranking page, they can click on a product page to see additional product information, read reviews and additional product information (Figure 2). Product pages feature the same shipping information, i.e. the badge, free shipping (if available) and the get-it-by date. After evaluating the alternatives, consumers have three options: (i) they can leave the website, (ii) they can purchase an item they clicked on, or (iii) they can continue searching. The process is illustrated in Figure A.19 .

If consumers purchase a product, then there are two ways in which the order can be fulfilled. First, the product can be fulfilled by the third-party seller as a Dropship product. This means that third-party seller sends the item to the consumer from its own warehouse. However, to facilitate shipping, in recent years, Wayfair has built multiple distribution centers that are located across the U.S., e.g., in California, Texas, and Georgia. These platform-owned warehouses are called Castlegate and create an alternative shipping option.[5] Third-party sellers on Wayfair have the option of storing their products in a Castlegate warehouse before the demand is realized.[6] The seller still owns the product but pays a rent to Wayfair to store the product in a Castlegate warehouse. Once the order arrives, the item is shipped directly from the Castlegate warehouse, and Wayfair handles all the shipping. Table 1 illustrates that while the Dropship method remains the primary fulfillment approach for the majority of orders (63%), there is a significant proportion of shipments (36%) accounted for by Castlegate or mixed Castlegate-shipped items.[7]

Products housed in Castlegate warehouses typically get fast shipping badge because the platform can ensure timely delivery to the consumer. However, Dropship sellers can also get the badge if

[4] Specifically, 2-day delivery radius of the warehouse where the product is located.

[5] See the Equity Research for the map of Castlegate warehouses.

[6] In Operations Research, this is called 'forward positioning' of inventory: sellers decide in advance how much inventory they are ready to allocate to a particular platform.

[7] 'Other' includes situations where the platform bought the inventory from the sellers for different business reasons and ships the products itself, but not necessarily from Castlegate.

Table 1 Summary statistics

Fulfillment Type	%Orders	Travel metrics			
		Miles	Weight	Girth	Locations
	(1)	(2)	(3)	(4)	(5)
Dropship	63.2%	1123.9	11.7%	5.0%	59,101
Castlegate	8.5%	793.8	7.5%	3.3%	25,894
Mixed	28.0%	936.8	11.5%	10.8%	17,950
Other	0.3%	852.6	11.7%	5.0%	8,637
Overall	100%	931.0	24.3%	10.7%	72,098

Notes. % Orders indicate the fraction of orders for each fulfillment type. Miles, Weight, Girth, Locations are corresponding metrics by fulfillment type.

(i) they can guarantee that the product will be delivered within promised number of days, 1 or 2 days, and (ii) if the seller maintains a good history of on-time delivery for fast badged products.

To summarize, fast delivery promises are prominent on the website and can potentially affect consumers' choices, especially if consumers have strong preference towards delivery speed. Regulators are concerned that the rush to fulfill fast delivery promises may result in inefficient shipping practices. At the same time, given platform's rules, most fast shipped products are located in platform-owned large distribution centers and this can lead to the economies of scale: platform sends millions of items in multiple locations and could be more efficient in consolidating the items compared to the situation when individual sellers send non-fast shipped items. In the sections that follow, we explore the causal effect of fast delivery promises on environmental emissions.

2.1. Data

In this subsection, we describe the main components of our data.The main sample consists of 1.5 months experimental data across all product categories. Unless stated otherwise, we focus on dining chairs category, which is a big product category on the platform.

Clickstream data. At the consumer-device-session-level, we observe full consumer browsing history on the website. Starting from the moment when consumers land on the website, we observe which categories they searched for and the ordered list of products they observed on the ranking page. For each product on each ranking page, we observe a product's price, ratings, numbers of ratings, and the availability of the fast shipping badge, the get-it-by dates and promised delivery days. Thus, we can fully re-create the view of the page as it was shown to the consumers. Next, we observe whether consumers clicked on an item, added it to cart or purchased.

Transactions and shipping data. At the consumer-level, we observe which products were ordered by a consumer, time of transaction, quantity, price, destination state, city, and zip code. From the clickstream data, we know whether the chosen product had fast delivery promises. We also observe which supplier was chosen to fulfill the order, and the address of a warehouse from which a product was shipped and consumers' destination address.

Shipment tracking data. We observe shipment tracking information provided by FedEx and UPS. Specifically, we observe the timestamp and origin scan of the shipment and all subsequent tracking scans up to the destination together with the timestamps. This data allows us to analyze whether multiple shipments were consolidated and shipped together from the warehouse or added en route to the same destination. Moreover, it allows us to credibly estimate how loaded a truck was which is crucial to analyze the environmental impact because the load of the truck may have significant implications on environmental emissions.

Supplier rejection logs. One of the key parts of the data for the counterfactuals are the records of suppliers that were rejected from fulfilling the order. We observe which supplier was selected to fulfill each historical order and the full ordered list of rejected suppliers. We use this data in the counterfactuals to credibly identify which supplier is most likely to fulfill the order from a specific Zip Code.

Product characteristics. We also use data on product characteristics, such as weight and girth of the products. This allows us to distinguish small, medium, and large parcel items. Specifically, the data contains information on product dimensions, and carton type used to ship the item.

2.2. Environmental emissions

In our calculation of environmental costs, we focus on CO2 emissions stemming from last-mile delivery, which refers to the emissions generated when an item is transported from a warehouse to the final consumer. Packaging costs are disregarded since Wayfair sells durable goods that are generally medium to large in size and necessitate separate packaging regardless of the shipping method used. While a retailer selling small items may be able to combine items like pens and hairpins in packaging, furniture items generally require separate packaging due to their size and nature.

Truck fuel emission costs. CO2 emissions vary based on the type of vehicle, and there are several online Carbon Footprint Calculators that determine the metric tons of CO2 emitted per mileage using vehicle characteristics such as vehicle manufacturing year, manufacturer, and model.[8] In our analysis, we adopt the methodology outlined by the Environmental Defense Fund (EDF), which incorporates three metrics - distance, weight, and emissions factor - to calculate CO2 emissions.[9]

$$\text{Greenhouse Gas Emissions from Shipping} = D \times W \times EF \tag{1}$$

where D is the distance the shipment has traveled (in miles); W is the weight (in pounds/tons) of shipment; EF is the constant emissions factor calculated by EDF. This approach was also used

[8] For example, refer to Carbon Footprint Calculator.

[9] See Green Freight Math: How to Calculate Emissions for a Truck Move

by Muñoz-Villamizar, Velázquez-Martínez, Haro, Ferrer and Mariño (2021). In Appendix B, we show that our results are robust regardless of whether we use miles traveled by a product, or miles multiplied by the product weight ($D \times W$). Given Equation 1, this means that all changes in the miles traveled translate proportionally to the CO2 emission changes. Therefore, unless stated otherwise, we utilize changes in miles traveled as a proxy for changes in CO2 emissions.

One might wonder why we do not account for electricity consumption in warehouses or for the emissions associated with shipping items from manufacturers to sellers' or Wayfair's warehouses. However, due to the design of the large-scale field experiment, described in Section 3, we are able to focus solely on last-mile delivery. This is because, operationally, the emissions related to warehouses and inbound shipments are already balanced and incurred in both the control and treatment groups. Therefore, we consider these emissions as sunk and balanced between the two groups, and we elaborate on this aspect further in the counterfactual section.

2.3. Descriptive statistics

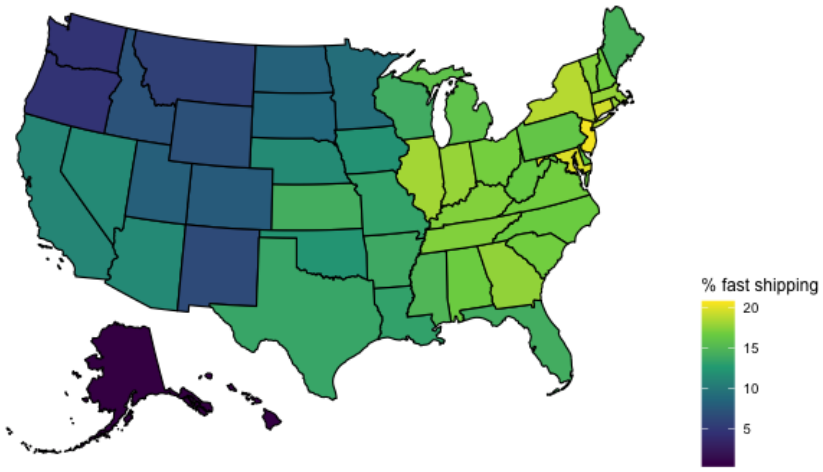
Before turning to the main analysis, we present some relevant summary statistics on the fast delivery promises. Fast delivery promises are contingent on the supplier being located within a 2-day delivery radius of the consumer's identified geolocation, as previously stated. Consequently, the penetration of fast delivery promises varies across different states. Figure 3 illustrates the percentage of products that have fast shipping badges among randomly sampled first five pages of the ranking page results. The regions such as the East Coast, West Coast, and Texas have the highest proportion of available fast shipping due to the concentration of numerous warehouses in those areas, both owned by the sellers and the platform. On average, the percentage of fast delivery promises in these regions falls between 10% and 20%, while regions like Hawaii or Alaska unsurprisingly have limited availability of fast shipping.

Products with fast delivery promises are usually placed higher in the ranking page results. Table 2 presents the distribution of average product ranks based on the type of product fulfillment. We categorized products into different groups based on their shipping options, including 1-day shipping, Free 1-day shipping, Free 2-day shipping, 2-day shipping, Free shipping with no delivery day promise, and products without any shipping flag. Column (1) shows that most observations (products) either have free shipping promise or do not have shipping flag, followed by free 2-day shipping and 1-day shipping promises.

Column (4) of Table 2 shows that products which do not have any shipping flag are typically shown on position 281, corresponding to page 5 or 6 of the ranking results.[10] At the same time,

[10] For this analysis, we only selected pages that had 48 products featured to make sure that the results are consistent.

Figure 3 Percent of fast shipping by states



products with free 1 or 2-day delivery promise are featured on positions 78 or 142, i.e. pages 2 or 4, respectively. Thus, products with delivery promises are featured at the top of the ranking results.

Notably, products with delivery promises tend to be slightly less expensive than those offering only free shipping or lacking any shipping flag, as indicated in Panel B of Table 2. For instance, Free 2-day shipped products have a median price of $228, while Free Shipping and No shipping flag products are priced at $269.99 and $249.99, respectively. This price discrepancy can be attributed to several factors. First, sellers who can afford to provide fast delivery options are often established and large sellers, which enables them to offer additional discounts or allowances on the wholesale price they provide to the platform. Since the platform applies a margin based on the seller-provided wholesale price,[11] these discounts are passed on to the retail prices. Second, there may be an endogeneity effect wherein products with fast delivery promises are featured in more prominent website locations, attracting greater demand and subsequently benefiting from economies of scale that result in lower shipping costs, among other factors.

Together with lower prices, fast shipped products benefit from more prominent product placement. Figure 4 demonstrates that the click-through rates are higher at the top of the ranking page (with lower product ranks being more favorable), gradually decreasing as the positions move down. The spikes observed at the 48th position are attributed to consumers being directed to a new page, where the top positions attract more clicks. The way ranking algorithms work is that

[11] Wayfair uses a dynamic pricing algorithm and takes wholesale cost and any discounts set by the sellers as an input in the pricing formula. See Wayfair pricing.

Table 2 Summary Statistics in Dining Chairs category

	Panel A: Average Product Rank					
	Observations (1)	Min (2)	Mean (3)	Median (4)	Max (5)	SD (6)
1-day shipping	4,005	1.00	193.72	116.34	1177.60	219.26
Free 1-day shipping	254	1.00	134.28	78.65	921.00	145.60
Free 2-day shipping	8,284	1.00	212.50	142.09	1190.80	231.06
2-day shipping	451	1.00	22.40	18.00	333.48	24.89
Free Shipping	66,024	1.00	165.18	29.00	1191.65	248.76
No shipping flag	24,780	1.00	362.13	281.50	1192.58	335.76
	Panel B: Price ($)					
	Observations (1)	Min (2)	Mean (3)	Median (4)	Max (5)	SD (6)
1-day shipping	4,005	12.99	254.71	225.83	1997.50	158.27
Free 1-day shipping	254	22.99	264.95	239.99	589.99	119.56
Free 2-day shipping	8,284	19.59	266.71	228.52	2699.99	187.51
2-day shipping	451	12.99	27.63	27.99	275.00	15.47
Free Shipping	66,024	13.08	410.25	269.99	2792.93	430.39
No shipping flag	24,780	12.99	319.20	249.99	2792.19	330.87

Notes: This table shows summary statistics by the product type.

products with larger number of clicks gain more prominence on the website, which may increase the purchase probability.

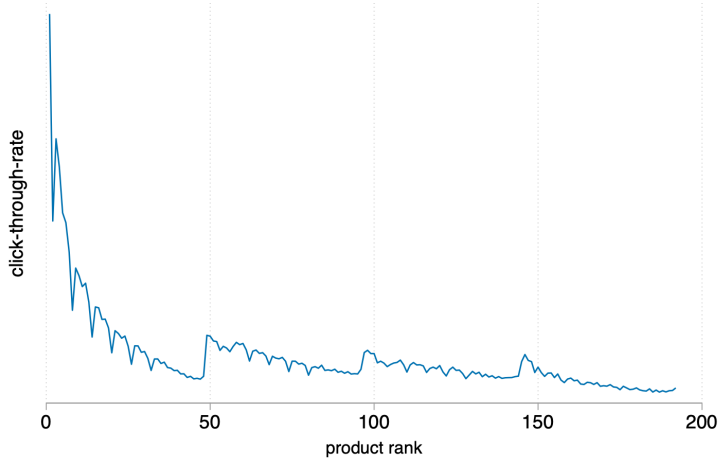
To summarize, products that have fast delivery promises are ranked higher on the ranking page results, they are clicked more frequently and set lower prices which can attract additional demand.

One of the key features of the fast delivery promises is that majority of them are fulfilled by the platform itself through the platform-owned Castlegate distribution system. This is also true for other platforms, such as Amazon. Amazon's widely known fast delivery program Prime also relies on the warehouses that are owned and maintained by Amazon itself.

The central feature of the platform-owned distribution centers is that they are strategically located closer to potential demand. Column (2) in Table 1 shows that the average miles traveled by a product through the Dropship (seller-based) system is 1,123 miles, while product sent via Castlegate travels 793 miles on average. This is a pure location benefit of the warehouses. In addition, there are economies of scale benefits from the centralized shipment strategy.

To quantify potential economies of scale, we used the UPS and FedEx shipment tracking information. For each warehouse from which at least one order has been shipped we calculated the

Figure 4 **Share of fast shipped products between control and treatment groups**



number of orders that are being shipped from that warehouse at a given point in time, which we measure at a minute level. We then split warehouses into platform-owned one versus third-party seller warehouses.

Figure 5 **Item consolidation across centralized distribution centers and decentralized seller system**



Figure 5 plots the number of orders that left a warehouse at a given point in time. We find that the median of 13 orders leave a platform warehouse, while 3.27 orders leave a single third-party seller warehouse at a given minute on a single truck. This shows that the platform-owned warehouses can lead to significant economies of scale because they carry more products that can be shipped together in one truck.

Thus, regulatory concerns about products that have fast delivery promises may be unsubstantiated given that these products are cheaper and can be consolidated during shipment.

However, it is an open question how consumer choices may affect environmental emissions in the absence of fast delivery promises. To answer that question, we now turn to the large-scale experiment ran by the platform.

3. Fast Shipping Experiment

To quantify the environmental impact of the fast delivery promises, we use the experiment ran on Wayfair. The experiment ran for 1.5 months from the mid-June till the end of July 2019 and spanned the entire website traffic.[12] During the experiment, 30% of traffic were assigned to the control group and saw business-as-usual product rankings and fast delivery promises. Treatment group, 70% of traffic, saw reduced number of fast delivery promises. Specifically, for each treated visitor, the platform (i) removed between 25% and 75% of the fast shipping badges, (ii) pushed back their get-it-by dates and (iii) removed the fast shipping sort boosts (Figures 6 and 7). Operationally, each treated user was randomly assigned an s-value where $s \in [25, 75]$ and s% of the eligible fast delivery promises would be removed from the consumer's ranking page. s-values were assigned at the beginning of the experiment and stayed constant during the entire experiment. Randomization checks at the consumer-level are provided in Table 11 .

Figure 6 Control condition

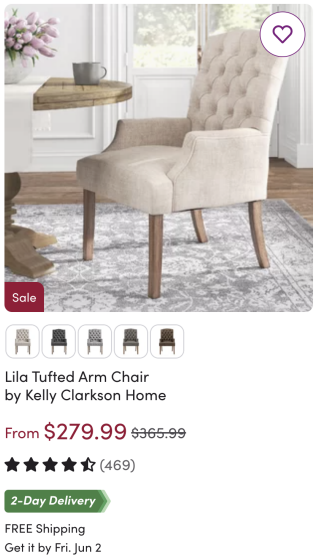
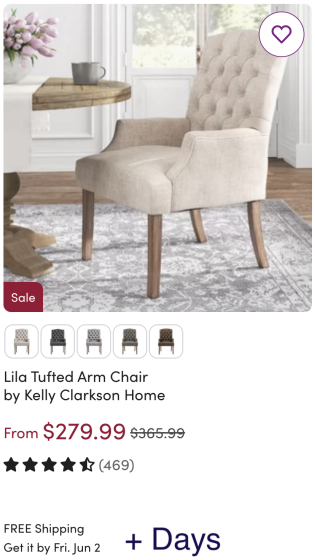


Figure 7 Treatment condition

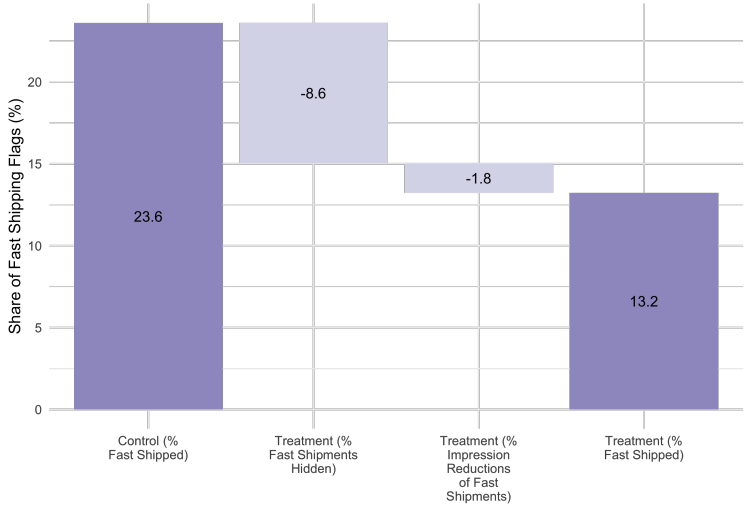


We validate that the experiment ran as intended by comparing the share of fast delivery promises on the ranking pages of the control and treatment groups. Figure 8 shows that in the control group approximately 24% of the products displayed to the consumers have fast shipping badges (first

[12] The platform paused the experiment during the Fourth of July sale.

bar).[13] In the treatment group, 8.6% of products that were supposed have fast shipping badge were displayed without one, as part of the experiment. Overall, treatment group had 13.2% fast shipping badges (last bar). We attribute the difference, 1.8% (=23.6-8.6-13.2)%, to the overall impression reduction of fast shipping badge-eligible products due to the fact that sort boost was disabled.[14]

Figure 8 Share of fast shipped products between control and treatment groups



Thus, treatment group consumers viewed 44% less fast shipping badges. Next, we validate the experiment by comparing the share of fast shipping badges hidden from the ranking page (Figure 10) and the share of fast shipping badges shown on the ranking page by s-value (Figure 11).

Figure 9 Experiment validation

Figure 10 % eligible fast shipping badges hidden by s-value

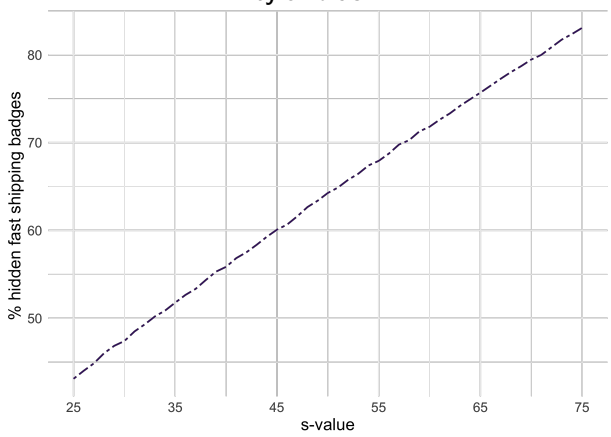
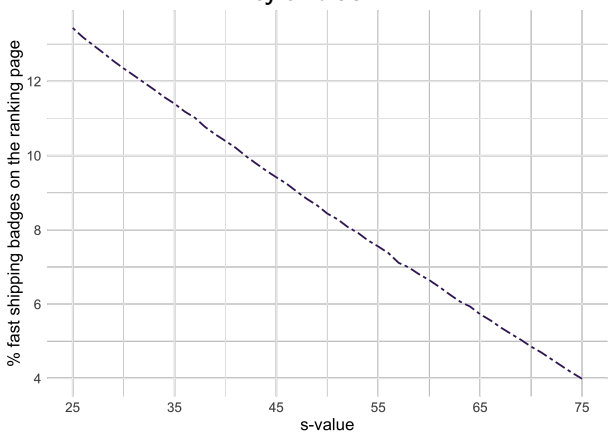


Figure 11 % products with fast shipping badges by s-value



[13] In the data, we observe not only the overall rankings, but the products that were loaded on the pages *viewed* by the customers.

[14] This is consistent with the platform's own analysis, and we thank Nathaniel Burbank for clear documentation.

Figure 10 illustrates a positive relationship between higher s-values and a greater percentage of hidden fast shipping-eligible products. Figure 11 shows that consumers with higher s-values saw lower percentage of fast-shipping badges on the page. Overall, the experiment ran as intended and changed the availability of the fast shipping badges on the website. However, it is important to note that the compliance is not perfect in the sense that consumers with an s-value of 45, for instance, had on average 60% instead of 45% of badges hidden.[15] Nevertheless, we will use s-value as an instrumental variable, whenever applicable, and interpret the results as the Intent-To-Treat estimates.

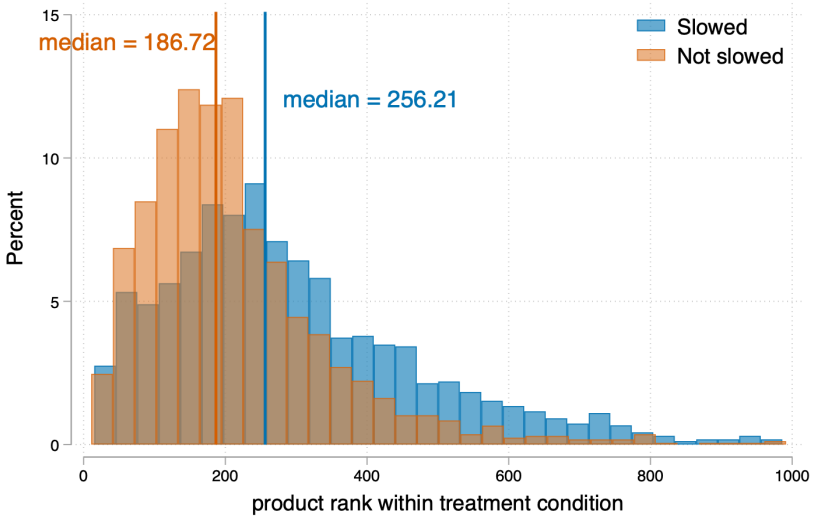
There are several important caveats associated with the experiment. First, the platform did not change the actual realized shipping speeds of the products for which fast delivery promises were removed. Due to the complexity of retail platforms' delivery network, changing the actual delivery times is hard to accomplish. One may be concerned that the results reflect not only the effect of removing fast delivery, but also the positive effect of getting delivery faster than initially expected. However, we think that this is not a big concern in our setting because we focus on the instantaneous outcomes, such as delivery time, prices paid, and abstain from looking at the long-term consumer demand effect. Moreover, although repeat purchases are common in dining chairs category, they usually occur 30-40 days after the initial delivery time, and given that the experiment lasted for one and half months only, we do not view it as a big issue.

Second, the experiment varied multiple aspects of product ranking pages: (i) it removed fast shipping badges, (ii) changed the get-it-by dates, and (iii) also removed sort boost which was standard absent the experiment. Thus, not only did the products got their badges removed, but they also were moved in the rankings in the treatment condition. In Figure 12, we illustrate the distribution of the product ranks in the treatment condition among *slowed* and *not slowed* products. Recall that different consumers had different s-values, which means that in the treatment condition a product could have had its fast shipping badge hidden for some pageloads and not for others. If the product had its fast shipping badge hidden, we call it *slowed* product and *not slowed*, otherwise. Products that were slowed in the treatment condition were ranked at position 256, while not slowed products in the treatment condition were ranked at position 186. Thus, indeed, products with hidden badges lost their prominence on the website, and were ranked 70 positions lower.

One could be concerned that we capture a combination of the ranking and the shipping effect in our results. However, the experiment actually replicates a real-life situation: had there been no fast shipping then there would be no product boosting too, and product rankings would change

[15] Conversations with the platform confirmed that the compliance was not perfect due to the technical complexity of the experiment.

Figure 12 Product ranks by treatment



considerably.[16] Thus, we view the treatment as the combination of removing the delivery promise and the corresponding ranking effect.

3.1. Experimental Results

Since this is a randomized experiment, we ran the following OLS specification to investigate how being randomly assigned to the treatment group affects different delivery and purchase outcomes:

$$y_i = \alpha + \beta treated_i + \varepsilon_i \tag{2}$$

where y_i is the outcome of interest for consumer i; $treated_i$ is the treatment assignment for consumer i equal to 1 in treatment group, and 0 in control; and ε_i is the error term.

The results in this subsection include all product categories and all experimental consumers who visited the ranking pages.[17] We take more than 47 million consumers who were part of the experiment, and estimate how treatment affected the number of orders, revenue and clicks. Table 3 shows that consumers in the treatment (reduced fast shipping) condition purchased less: number of orders decreased by 0.4% (Column 1) and revenue decreased by 1.5% (Column 2). However, we do not find significant differences in the number of clicks or the probability of clicking between treatment and control groups (Columns 3 and 4). Thus, consumers seemed to have clicked on the products as usual, but treatment group consumers purchased less. The latter could happen either because of the higher prices (see Table 2), or because consumers were unsatisfied with the delivery times.

[16] On Amazon, for instance, Prime products are ranked higher, and consumers have the ability to refine search results to only see Prime products. Had Prime been removed, that would also change the product rankings.

[17] We excluded consumers who arrived through the App because they typically see different rankings.

Table 3 Experimental results: consumer purchase outcomes

	(1) log(# of orders)	(2) log(revenue)	(3) log(# of clicks)	(4) clicked
Treated = Reduced Fast Shipping	-0.004***	-0.015***	0.000	-0.000
	(0.000)	(0.000)	(0.000)	(0.000)
Intercept	0.066***	0.300***	0.460***	0.512***
	(0.000)	(0.000)	(0.000)	(0.000)
Observations	47,299,888	47,299,888	47,299,888	47,299,888

Notes. This table reports the output from the estimation of equation 3 on **all experimental consumers**. Data is at the consumer-level. Columns (1)-(4) use OLS specification results. Robust standard errors in parentheses. Statistical significance: $^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$.

Next, we investigate how treatment assignment impacts the probability of purchasing products that can actually be shipped fast. We estimate the following specification with the consumer-purchased product level data:

$$y_{ij} = \alpha + \beta treated_i + \varepsilon_{ij} \tag{3}$$

where i is a subscript for the consumer, y_{ij} is an outcome variable of interest for consumer at a consumer-purchased product level ij; $treated_i$ is the treatment assignment of consumer i; ε_{ij} is the error term.

Table 4 Experimental results: consumer purchase outcomes

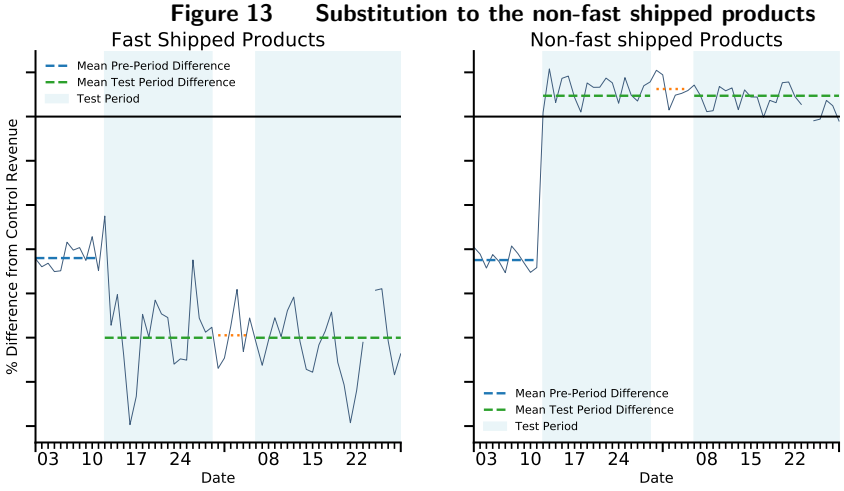
	Logit	*OLS*	
	(1) fast shipped	(2) log(distance) (miles)	(3) distance (miles)
Treated = Reduced Fast Shipping	-0.069***	0.010***	6.086***
	(0.001)	(0.001)	(0.425)
Intercept	-1.037***	6.400***	934.999***
	(0.001)	(0.000)	(0.215)
Observations	15,676,935	15,676,935	15,676,935

Notes. This table reports the output from the estimation of equation 3 on **all purchasing consumers**. Data is at the consumer-purchased product level. Column (1) uses the logit specification and Columns (2)-(3) use OLS specification results. Robust standard errors in parentheses. Statistical significance: $^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$.

Note that consumers in the treatment group do not know pre-purchase that the product will be shipped fast even if it does not have a fast shipping badge. Column (1) of Table 4 shows that consumers in the treated group are 6.9% less likely to purchase a product eligible for fast shipping. This result highlights that consumers in the treatment group substitute away to the non-fast shipped products and purchase them instead of fast shipped products. This happens because consumers see reduced number of fast shipping badges and the product ranking change overall.

Next, we investigate whether given consumers' choices one can conclude that removing fast shipping badges was an environmentally-friendly decision. However, Columns (2) and (3) of Table 4 show that in the reduced fast shipping condition, the average distance traveled by a product was 1% higher than in the control group, or 6 miles more. Thus, removing fast delivery promises led to a substitution towards products from more distant warehouses. Figure 13 confirms this.

Figure 13 Substitution to the non-fast shipped products



Notes. The y-axis is hidden for data sensitivity reasons.

On the x-axis of Figure 13, we plot the dates in July 2019, when the experiment was ran. The y-axis of the left-hand figure shows the % difference in revenue of fast shipped products compared to the non-fast shipped products before and during the experiment. The solid black line is zero and the shaded blue region is the experimental period with the gap for July 4th sale, when the experiment was paused. Thus, before the experiment, revenue of fast shipped products was lower than that of non-fast shipped products. This is intuitive because there are less number of fast shipping-eligible products than the non-badged products (recall Table 2). During the course of the experiment, the percent difference in revenue decreases, suggesting a decrease in the number of fast-shipped products sold. This is consistent with the result obtained in Column (1) of Table 4 that the probability of purchasing a fast shipping product decreases by almost 7%. On the right-hand side figure, we plot the mirror measure: the % difference in the non-fast shipped products' revenue compared to the fast shipped revenue. We normalized the difference to match the before-experimental period on the left-hand side figure. Unsurprisingly, the revenue of non-fast shipped products shows an immediate increase once the experiment commences. The

effects persisted even during the July 4th period, when the experiment was turned off. This could be explained by the persistence of the product ranking effects.[18]

Thus, in the experimental condition, there was a significant substitution towards non-fast shipped products that are typically located further away from the consumers. This led to 1% higher miles traveled. Based on the explanations in Section 2, we can conclude that the increase in miles traveled is proportional to the increase in CO2 emissions.

However, we also have to take into account additional environmental emissions stemming from the platform's potential inability to consolidate the orders under insufficient demand. Table 5 shows the back-of-the-envelope calculations. In the first row, we calculate the total number of miles-ton traveled before the experiment. In Table 1, we show that a product from the platform-owned warehouse travels approximately 793 miles and the Dropship (seller-owned product) travels 1,123 miles. Relying on publicly available sources, we assume that the average truck load for a full truck is 1.3 tons.[19] Next, in Figure 5, we showed that the median number of orders leaving a dropship (seller-owned) warehouse is 75% less than that of platform-owned warehouse (3.27/13.24). Back-of-the envelope calculations would suggest that the trucks are 75% loaded for the sellers compared to the platform. Combining all of this, a single truck would travel 2,125 miles-ton before the experiment.

Table 5 Back-of-the envelope calculation of economies of scale

	Platform-owned		Seller-owned		Total
	Miles traveled	Weight	Miles traveled	Weight	
Before experiment	793 miles[1]	1.3 tons[3]	1,123 miles[1]	1.3×0.75 tons	2,125 miles-ton
Savings during experiment	-793 miles	1.3×0.17 tons[4]	1,123 miles	1.3×0.75×0.21 tons	54.6 miles-ton
% Difference					+2.5%

[1] Taken from Table 2 Column (1) - Castlegate
[2] Taken from Table 2 Column (1) - Dropship
[3] Average estimated load for a truck - link
[4] Platform-owned warehouses lose approximately 17% of demand and the seller-owned products get 21% more demand compared to the historical numbers.

Compared to the pre-experimental period, during the experiment, platform-owned warehouses lost approximately 17% of orders and non-fast shipped sellers gained 21% of orders compared to the historical numbers. We calculate this by running difference-in-differences specification before

[18] In 2019, the platform employed a simple algorithm to rank products, where the algorithm only accounted for historical product clicks and purchases. Thus, if the non-fast shipped products gained more clicks they would keep their prominence even when the experiment was turned off.

[19] See link.

and during the experiment at the product level. Specifically, we estimate the following regression at the product level:

$$y_j = \alpha + \beta_1 fast_j + \beta_2 slowed_j + \beta_3 treated_j \times slowed_j + \varepsilon_j \quad (4)$$

where y_j is the number of orders of product j, $slowed_j$ is equal to 1 if product j was slowed and 0, otherwise, $treated_j$ is the indicator for whether the product was in the treatment group during the experiment. Table 6 shows that products that had slowed shipment lost 17.4 - 18.6% of the orders.

Table 6 Effect of fast shipping on the orders

	Number of orders (Poisson regression)		
	(1) overall	(2) 1-day badge	(3) 2-day badge
Slowed	-0.052***	-0.052***	-0.052***
	(0.003)	(0.006)	(0.004)
Treated	0.413***	0.519***	0.366***
	(0.003)	(0.005)	(0.003)
Treated× Slowed	-0.178***	-0.186***	-0.174***
	(0.004)	(0.007)	(0.005)
Intercept	0.554***	0.368***	0.641***
	(0.002)	(0.004)	(0.002)
Observations	528,836	184,388	344,448

Notes. This table reports the output from the estimation of equation (4). Data is at the product level. Robust standard errors in parentheses. Statistical significance: $^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$.

As a result, there were environmental savings on the platform side because there was a decrease by 793×1.3×0.17 miles-ton. However, since other sellers got higher demand and they travel on average larger number of distances, this leads to the additional 1,123×1.3×0.75×0.21 miles-ton traveled. Thus, during the experiment there was an increase of 54.6 mile-tons of truck travel distance. Using Equation 1 and assuming a publicly available number of 161.8 grams of CO2 per mile-ton, the reduction in fast shipping badges led to 2.5% increase in CO2 emissions or 8.8 thousand additional CO2 grams emitted per trip.

Thus, reduction in fast shipping badges led to (i) 1% increase in CO2 emissions because products were shipped from more distant warehouses, and (ii) an additional 2.5% increase in CO2 emissions due to loss in the economies of scale that centralized distribution centers can provide.

Finally, we investigate how consumers' product choices translated to delivery times and price differences. Table 7 shows the results of estimating specification 3 with delivery time, log price and log shipping cost as outcome variables. Column (1) shows that the wait time increased by 2.4%

or 0.133 days (=3 hours) for the consumers in the treatment group. Moreover, they bought 4.3% higher priced items. Recall from Table 2 that the non-fast shipped products are also more expensive than the fast shipped ones. Moreover, Column (3) shows that the shipping costs increased by 2.8%.

Table 7 Experimental results: consumer delivery outcomes

	OLS		
	(1) delivery time	(2) log(price)	(3) log(shipping cost)
Reduced Fast Shipping	0.133***	0.043***	0.028***
	(0.003)	(0.001)	(0.001)
Intercept	5.511***	4.283***	2.560***
	(0.001)	(0.000)	(0.000)
Observations	15,593,603	15,676,935	15,676,935

Notes. This table reports the output from the estimation of equation 3 on **all purchasing consumers**. Data is at the consumer-purchased product level. Robust standard errors in parentheses. Statistical significance: $^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$.

These results show that not only did the removal of fast shipping promises increased the CO2 emissions, but it also worsened consumer outcomes, both in terms of the delivery wait time and the prices paid for the products. Thus, despite the regulatory concerns over fast delivery promises, mere removal of the fast delivery promises may not be a desirable solution. Instead, in the next sections, we propose and discuss an alternative solution that could help platforms reduce environmental emissions.

4. Alternative Ranking Algorithm and Counterfactuals

In the previous section, we established that removing fast delivery promises hurts consumers and increases CO2 emissions. Therefore, we propose and evaluate alternative solutions to help platforms reduce environmental emissions. This is a multi-objective optimization problem. On the one hand, platforms have to maintain customer satisfaction and the conversion (purchase) probability. On the other hand, there is an objective to decrease environmental emissions given that there are explicit Environmental, Social, and Governance (ESG) standards[20] for retail platforms. Moreover, note that retail platforms' incentives may be aligned with the environmental goals because less emissions mean lower shipping costs. Thus, the overall goal is to reduce the emissions and shipping costs without compromising user experience. We start by describing the static algorithm and will afterwards explain a more complicated dynamic algorithm that could solve this problem.

[20] See the example standards.

4.1. Static algorithm

Formally, multi-objective optimization problem could be written in a simple static way as follows:

$$\max_{\eta} \eta \cdot P(conversion) - (1-\eta) \cdot CO2\ Emissions \tag{5}$$

where $P(conversion)$ is the purchase probability, $CO2\ Emissions$ are the carbon footprint from shipping, and η is the weight that the platform assigns to the conversion goal.

The objective is to choose optimal weight η. One way to solve this problem is as follows. First, the platform creates a set of popular products that have high purchase probability. The platform already incorporates this goal in the current product ranking algorithm, so it knows which products have high purchase probability. Second, the platform has to decide which products will incur less CO2 emissions. A simple heuristic we adopt is that products physically closer to the consumer are the ones with less CO2 emissions.[21] Thus, for each consumer who arrives to the website, the platform has to identify the set of products that are located closer to the consumer. Finally, the platform decides between showing products with higher conversion probability and the ones with lower emissions.[22]

We call this algorithm static because we assume that the platform solves the problem once either through simulations or experimentation, and this formulation does not allow for learning. Next, we discuss a more complicated multi-armed based approach that can help platforms update their beliefs as they explore potential solutions.

4.2. Dynamic algorithm

Ideally, a good algorithm would show either high conversion probability or low CO2 emissions products to each consumer and will learn from consumers' click and purchase behavior whether they like the products. One could formulate this as a multi-armed bandit algorithm. The algorithm chooses a product (an arm) or a set products (rankings) to show to the consumer at the first iteration (trial). Next, if consumer clicks on a product, then the product receives a payoff of 1 and 0, otherwise. As more and more consumers interact with the products, the algorithm updates its posterior beliefs about products' attractiveness. The algorithm eventually learns which products are the best to show.

There are several challenges with standard multi-armed bandit algorithms. First, a large retail platform carries millions of products and hosts millions of consumers every day. An algorithm

[21] One could think about the potential economies of scale from shipping from the same warehouse as was discussed in the previous section. However, our goal in this part of the paper is to propose the simplest possible algorithm that any platform can use.

[22] Note that we assume away any product personalization that the platform can employ apart from CO2 emissions-related personalization.

should be computationally feasible, which is hard to accomplish with millions of products. Another challenge is that we need to optimize two goals: conversions and lower emissions. Different regions have different assortment of products in the nearby warehouses. Therefore, different products should be shown in different geographical regions, which further complicates the problem. To address these challenges we adopt the contextual bandit algorithm from Li, Chu, Langford and Schapire (2010). First, we describe the algorithm itself, and then explain how we apply it to our problem.

At time t, there is a consumer i_t and a set of products (arms) A_t that can potentially be shown to the consumer. There is also a feature vector $x_{a,t}$ that summarizes information about the user i_t and a product (arm) $a \in A_t$. For instance, $x_{a,t}$ could be a vector that contains a product a's price, conversion probability, distance from the focal consumer; or the state in which consumer lives. At iteration t an algorithm shows a product a_t and observes a payoff of r_{t,a_t} from consumer i_t. Suppose that the payoff is one if consumers clicks on a presented product, and zero, otherwise. If consumers clicks on a product, the algorithm infers that the consumer likes the product, and vice-versa. Collecting information on clicks/no-clicks allows the algorithm to learn which products consumers like.

The total expected payoff for T iterations is $E[\sum_{t=1}^{T} r_{t,a_t}]$ and the optimal expected payoff is $E[\sum_{t=1}^{T} r_{t,a_t^*}]$, where a_t^* is the product (or set of products) with the maximum expected payoff. In bandit literature, the common formulation is to minimize the regret from playing a wrong arm (showing zero payoff product) with respect to the optimal product selection. The T-iteration regret $R_A(T)$ of the algorithm is defined as the difference between the optimal expected payoff and the overall expected payoff:

$$R_A(T) = E[\sum_{t=1}^{T} r_{t,a_t^*}] - E[\sum_{t=1}^{T} r_{t,a_t}] \tag{6}$$

Following Li, Chu, Langford and Schapire (2010), we use the linear UCB algorithm to generate rankings. The main idea is to formulate the expected payoff in a linear form, which in turn leads to a computationally tractable solution for the problem. The expected payoff conditional on context vector $x_{t,a}$ can be written as:

$$E[r_{t,a}|x_{t,a}] = x_{t,a}^T \theta_a^* \tag{7}$$

where $x_{t,a}$ is the $d \times 1$ vector and θ_a^* is the $d \times 1$ vector of unknown coefficients. The d-dimensional context vector can include both consumer and product-related information. In our application, one could think of a context vector as the state in which consumer lives, historical purchases, environmental savings from the product shipments etc. The goal is to recover unknown parameters θ_a^*.

Suppose that in the data we observe a design matrix D_a with the dimension $m \times d$ where m is the number of contexts available for a product a. Assuming a one-dimensional context vector where the only element is the state in which consumer lives, one could think of D_a as a $m \times 1$ vector for each product, where $m = 50$ for fifty US states.[23] Next, from the historical data, we observe a vector $c_a \in \mathbb{R}^m$ that would include a binary payoff, either click or no click for each context (state) for each product a. Generalizing 7 to multiple contexts, we could write the problem as

$$c_a = D_a^T \boldsymbol{\theta_a} \tag{8}$$

One could write the solution to this problem in the form of Ridge or Lasso regression estimate:

$$\boldsymbol{\theta_a} = (D_a^T D_a + I_d)^{-1} D_a^T c_a \tag{9}$$

where I_d is the $d \times d$ is the identity matrix. Given this general solution and applying UCB algorithm for optimal product search, at each iteration t, the algorithm chooses

$$a_t = \underset{a \in A_t}{\arg\max}(x_{t,a}^T \hat{\theta}_a + \alpha \sqrt{x_{t,a}^T A_a^{-1} x_{t,a}}) \tag{10}$$

where α is an exploration rate. Thus, at each iteration, the algorithm exploits the product space by choosing a product with highest expected payoff, $x_{t,a}^T \hat{\theta}_a$. At the same time, the algorithm explores product space by showing products with uncertain payoff, $x_{t,a}^T A_a^{-1} x_{t,a}$. The level of exploration is determined by parameter α. If α is high, then algorithm tends to try products with uncertain payoffs and the reverse is true for the low α. Li, Chu, Langford and Schapire (2010) prove that the regret bounds on this algorithm are finite, which means that the solution is near-optimal.

There are still several challenges with applying the algorithm as-is in our application. First, the platform carries more than 50,000 products in dining chairs category which makes it computationally hard to optimize over all products. Next, we need to generate the *rankings*, an ordered list of multiple possible products, by choosing between products with high conversion probability and low CO2 emissions.

We adapt the algorithm for our setting as follows. The context vector consists of a state in which consumer lives, implied CO2 emissions from shipping a product to a state, product's conversion probability, price, consumer's historical purchases of products in a given category, and consumer's demographic information. The resulting context vector is an eight-dimensional vector. CO2 emissions, product's purchase probability and price are continuous variables, which is why we bin them to have finite number of contexts. We further truncate the number of contexts by setting

[23] We combine DC with Virginia.

an upper bound on possible shipping costs of a product.[24] This reduces the number of products available at each trial making the problem computationally feasible.

Thus, at each iteration for a consumer in state s, the algorithm chooses a product among the available set of products. To get the rankings, we generate 72 best products for each consumer by iteratively removing the best product and re-optimizing the problem again.[25] We evaluate the algorithm by simulating consumer search and purchase outcomes via search model. Since we use the search model for the evaluation of both static and dynamic algorithm, the next section describes the model and the estimation results.

5. Model

To evaluate the alternative product ranking policies and speak to the changes in consumer welfare, we estimate a sequential search model. The model was proposed by Weitzman (1979) and has been widely used in Marketing literature (see Kim, Albuquerque and Bronnenberg 2010, Ursu 2018).

Suppose there is a consumer who searches Wayfair website for a dining chair. First, consumer is directed to the ranking page that features J products in a matrix form. Wayfair typically displays fourty eight products on each page, and the products are arranged into 12×4 matrix.[26] On the ranking page, consumer already observes some information about products, such as price, rating, number of ratings, image, fast shipping badge and delivery promise in days. Upon clicking on a product on the ranking page, consumer will reveal additional product information, such as detailed product specifications and product reviews. Finally, consumers decide between continuing to search, purchasing or leaving the website.

Following Kim, Albuquerque and Bronnenberg 2010, we formalize the problem as follows. On the ranking page, a consumer has some expected utility v_j about product j's utility prior to clicking. Consumer has to search, i.e. click on the product, to reveal the true product utility u_j, which we write as $u_j = v_j + \varepsilon_j$. Thus, upon clicking on a product, consumer reveals ε_j which resolves the uncertainty about the utility u_j. We assume that ε_j is a random variable drawn from normal distribution with zero mean and product-specific uncertainty σ_j.

Consumers have to pay a search cost to reveal ε_j. One could think of search costs as opportunity costs of time spent on the website or mental cost of processing the information on the product page. In the data, we observe that consumers are more likely to click on the items that are located

[24] We cannot reveal the exact upper bound due to data sensitivity, but any platform could set a reasonable upper bound based on their own shipping costs.

[25] The linear structure of the problem allows for efficient computation.

[26] The layout might be different depending on the screen size and consumers' device settings, but most often consumers see four products per row.

at the top of the ranking pages (lower positions), which is why we model search costs as a function of product positions (ranks). Following Chen and Yao (2017) and Ursu (2018), the search cost function should satisfy $c_j(p_j) > 0$ (search costs are positive) and $c_j'(p_j) > 0$ (margnal search costs are increasing with position).

Suppose a consumer clicked on some products. Next, she has to decide whether to continue searching or stop. If she decides to continue searching, she has to choose which product to click on. If she decides to stop the search process, she has to decide whether to purchase the best product searched so far or to leave and collect outside option utility, which is normalized to zero. Technically, this is a complex dynamic problem, in which consumer keeps track of the best utility observed so far ($\hat{u}$) and the set of unsearched products ($\tilde{J}$), $S = (\hat{u}, \tilde{J})$. Weitzman (1979) completely characterizes a closed-form solution to the dynamic problem. This solution relies on several important assumptions.

Assumption 1 **(No learning)**. *We assume that consumers do not update their beliefs about other products after clicking on a product. That is, $v_k,\ k \neq j$ do not change after clicking on a product.*[27]

Assumption 2 **(Unit demand)**. *We assume that consumers have unit demand and do not repeat purchase.*[28]

According to Weitzman (1979) solution, each product j is assigned a reservation utility z_j, which is defined as the critical number that makes consumer indifferent between searching product j or not. Formally, reservation utility of product j is computed as the value that equates the marginal cost of searching and marginal utility from searching:

$$c_j = \int_{z_j}^{\infty} (u_j - z_j) f(u_j) du_j \tag{11}$$

Recall that consumer's state space consists of the best utility observed so far and the set of unsearched products. Weitzman (1979) shows that the optimal search process of the consumer under some assumptions, can be characterized by three simple rules. At each point in time, consumer chooses the highest reservation utility product if she decides to click on a product [**selection rule**]. She decides to stop if the maximum utility observed so far exceeds the reservation utility of any unsearched product [**stopping rule**]. When consumer chooses to stop, she will choose the option with the highest utility observed so far. Note that the outside option is always searched

[27] For search with learning, see Hodgson and Lewis (2022), Santos, Hortaçsu and Wildenbeest (2017).

[28] In dining chairs, repeat purchases are not uncommon, however, most consumers typically repeat visit only a month or longer after the first purchase. Given that the experiment ran for one and half months, we rule out repeat purchases.

and the utility is set to zero. Following Kim, Albuquerque and Bronnenberg (2010), one can solve for the reservation utility as follows

$$\begin{aligned} c_j &= \int_{z_j}^{\infty} (u_j - z_j) f(u_j) du_j \\ &= \left((1 - \Phi(\frac{z_j - v_j}{\sigma_j})) \right) \left(v_j - z_j + \sigma_j \frac{\phi(\frac{z_j - v_j}{\sigma_j})}{1 - \Phi(\frac{z_j - v_j}{\sigma_j})} \right) \end{aligned}$$

The closed-form solution for the reservation utilities is

$$z_j = v_j + \sigma_j m_j \tag{12}$$

where m_j solves $\frac{c_j}{\sigma_j} = (1 - \Phi(m_j))(\lambda(m_j) - m_j)$ where $\lambda(\cdot)$ is the hazard rate function.

Given the closed-form solution for z_j one can estimate the model by maximizing the likelihood of observing the search paths that are present in the data.

Estimation. According to the model, first, consumer has to order products in the descending order of their reservation utilities. Thus, products are ordered as $R_i = [j(1), j(2), ..., j(N)]$ where $j(1)$ is the product with the highest reservation utility, $j(2)$ is the product with the second highest reservation utility and so on.

Selection rule requires that consumers first click on a product with the highest reservation utility. All subsequent clicks, if any, should also feature products with highest reservation utilities. Thus, product searched at kth search, $j(k)$ should have the highest reservation utility among all remaining unsearched products.[29]

$$z_{j(k)} \geq \max_{l=k+1}^{J} z_{j(l)}, \ \forall l = 1, ..., J-1 \tag{13}$$

Stopping rule requires that the consumer's reservation utilities for the unsearched products should be lower than the maximum utility observed so far, say at search iteration l, $\hat{u}_l$.

$$z_{j(l)} \leq \hat{u}_l \tag{14}$$

Moreover, stopping rule requires that if consumer searched product l, then the reservation utility from product l should be higher than all the utilities observed so far.

Choice rule requires that the consumer's chosen utility (e.g. product j's utility) is the highest among all.

$$u_j \geq \max u_l, \forall l \in searched \tag{15}$$

[29] We drop consumer subscript for convenience.

We can then write the probability that consumer i searches in some order R_i and chooses product j as

$$\begin{aligned} P_i(j) &= P(selection, stopping, choice) \\ &= P\Big(z_{j(k)} \geq \max_{l=k+1}^{J} z_{j(l)},\ \forall l = 1, ..., J-1 \\ &\quad \cap\ z_{j(l)} \leq \hat{u}_l \\ &\quad \cap\ u_j \geq \max u_l, \forall l \in R_i \Big) \end{aligned}$$

We then estimate the underlying preference parameters by maximizing the likelihood of observing the data, where for each consumer the log-likelihood is:

$$LL_i = \sum_{R_i} \sum_{j} d_{jR_i} log P(j)$$

where R_i is the search order as defined above, and d_{jR_i} is equal to 1 if consumer chooses search order R_i and product j. The total log-likelihood sums Equation 5 over all consumers i. We follow the estimation procedure as described in Ursu (2018) using the simulated log-likelihood approach.

5.1. Estimation Results

Before estimating the model on real data, we confirm that the model can recover the true parameters via Monte-Carlo simulations. We simulate the search process described above for 1,000 and 10,000 consumers with the same set of true parameters. Table 8 shows that the sequential search model recovers both true preference and true position effects well, which ensures that our estimation procedure will deliver credible estimates.

Table 8 Monte-Carlo Simulation Results

	True values	Estimates	
		$N = 1,000$	$N = 10,000$
Price	-2.5	-2.612	-2.502
		(0.213)	(0.073)
Attribute 1	1.2	1.327	1.264
		(0.039)	(0.023)
Attribute 2	0.5	0.713	0.662
		(0.492)	(0.196)
Constant	-4.5	-4.219	-4.245
		(0.001)	(0.000)
Position effect	0.5	0.511	0.533
		(0.001)	(0.000)
Log-likelihood		1,581	16,257

Next, we take the model to the real data and estimate the parameters of interest. The parameters of the utility function are preferences β's over x_j=[price, rating, free shipping, 1-day delivery, 2-day

Table 9 Estimation Results

	Estimates	
	$\hat{\beta}$	*st.err.*
Price ($)	-0.430	(0.008)
Rating	0.125	(0.019)
Free Shipping	0.127	(0.006)
1-day delivery	0.097	(0.000)
2-day delivery	0.107	(0.001)
Sale/Promotion (%)	0.067	(0.003)
# Ratings	0.003	(0.001)
Brand	0.014	(0.003)
Search cost	-0.045	(0.003)
Position effect	0.034	(0.001)
Log-likelihood	7,933.2	
# Consumers	15,000	

delivery, price discount %, # ratings, large brand dummy]. Moreover, we have to estimate search costs.

Table 9 shows the estimation results for 15,000 consumers. The signs on the coefficients make sense: consumers dislike higher prices (-0.430 coefficient), they like high rating, free shipping, 1 day delivery and 2-day delivery. They are not as sensitive to discounts, number of ratings or large brand dummies. This is intuitive given that dining chairs do not have a major brand names that consumers would know, as would be the case for the video cameras, for examples. Search costs and position effects are estimated to be $0.10 (0.045/0.430) and $0.07 (0.034/0.430).

Given these parameter estimates, we can now run counterfactual simulations to understand how consumers' search and purchase decisions will change under different sets of rankings.

6. Counterfactuals

We evaluate both static and dynamic algorithms as follows. First, we generate the counterfactual rankings using each algorithm. Next, we fix estimated search model parameters and simulate consumer search and purchase outcomes under the counterfactual rankings. Finally, we calculate the resulting shipping distance and shipping costs using the historical supplier rejection logs. The data is available starting 2018 and provides a good proxy for which supplier would be selected when a product is ordered from a particular location. Given that we show the same rankings at the city or state level, we have 99% coverage of potential shipment routes. This allows us to calculate the counterfactual shipping distance and shipping costs. In the dynamic algorithm, we encounter an additional issue of products potentially going out of stock because the algorithm

keeps recommending good product to each consumer. We account for that by setting a manual rule for each product given its historical inventory levels available.[30]

6.1. Static Algorithm Results

First, we run a counterfactual evaluating the static algorithm. Recall that the problem is formulated as follows:

$$\max_{\eta} \eta \cdot P(conversion) - (1-\eta) \cdot CO2\ Emissions \tag{16}$$

where the objective is to choose optimal η. We run simulations as follows. We bin η in a grid of $[0,1]$ with an interval of 0.1. Suppose we set η to 0.5. This means that equal priority is put on both conversion and emission goals. Therefore, for each consumer we generate a set of rankings of 72 products where 50% share of products are the ones with high conversion probability and 50% of products are the products physically closer to each consumer. Fixing estimated search model parameters, we simulate how consumers respond to the new rankings.

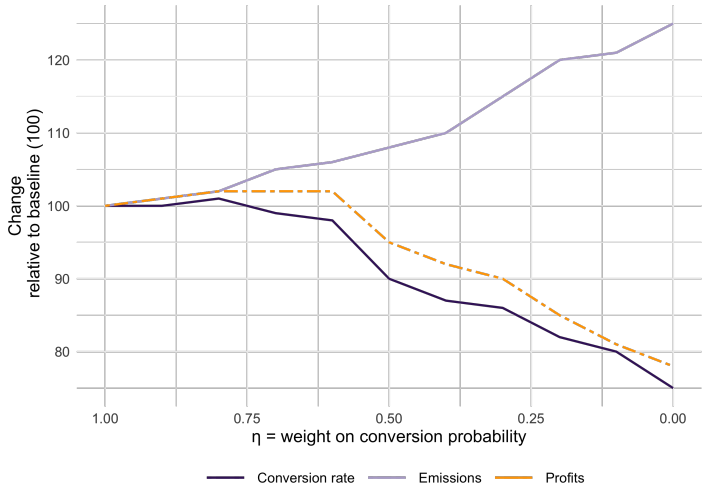
Due to the static nature of the problem, it is fairly easy to generate the rankings for a simulated set of consumers. Note that the set of best conversion probability products from which we sample is same for all consumers. However, the set of environmentally-friendly products is different because consumers are located in different regions. During simulations, we assume the same distribution of geographical locations from where the traffic arrives. In this counterfactual, the region is the city/town consumer is located in.

Figure 14 shows the results of the counterfactuals where we set the parameter η to 1 and proceed with the step of -0.1 up to 0. Due to the data confidentiality restrictions, the y-axis in the figure is the percent change compared to the baseline of $\eta = 1$ - full emphasis on conversion probability. The dark purple line shows the changes in the conversion rates, the light purple line shows the emissions savings which we calculate in the same fashion as in Section 2, and dashed yellow line shows the profit changes.

At $\eta = 1$ all three lines are set at 100 for benchmark comparison. As we decrease η, i.e., put more emphasis on the environmental objective, conversion rates start decreasing with the drastic decrease starting at $\eta = 0.6$. Clearly, the environmental savings start to increase. An important insight from this counterfactual is that profits are maintained even at a higher than baseline level up to $\eta = 0.6$, after which they start to drop. The reason profits are sufficiently high at $\eta < 1$ is that even if conversions decrease, the shipping costs for the products also decrease. Therefore, the

[30] It is a simple constraint, where we switch to the next best seller if the number of orders of a product start exceeding average historical inventory levels.

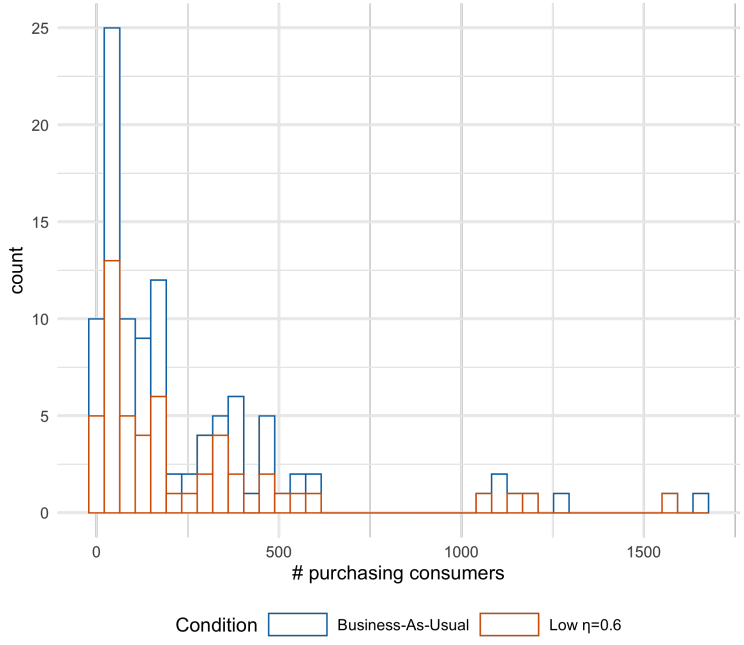
Figure 14 Counterfactual Results



profit margin still increases. However, after $\eta = 0.6$ the platform's profit metric start to decrease, which is clearly an undesirable outcome for the platform.

At $\eta = 0.6$, although conversions decrease the distribution of purchasing consumers across states does not change as much as is illustrated in Figure 15.

Figure 15 Number of purchasing consumers across states



Notes. This figure illustrates the distribution of the number of purchasing consumers across different states in the business-as-usual (control) condition and the low $\eta = 0.6$ condition.

These counterfactuals highlight that there may be an optimal value of η that can serve both the conversion and the environmental goal and even lead to higher profit margins. Thus, retail platforms could experiment by alternating between bestseller and environmentally-friendly products in the product rankings.

One of the challenges with this approach, as was mentioned above, is that algorithm is static and does not learn and cannot react to changes in consumer preferences or to the overall feedback of the consumers. Extensive experimentation with different η's is possible but is costly. Therefore, next we evaluate a dynamic algorithm that is reactive to consumer preferences.

6.2. Dynamic Algorithm Results

We illustrate how the counterfactuals work with a simple example. Suppose there are only five contexts (e.g., five states) and eight possible products $1, ..., 8$, of which products 3, 4, 6 and 8 are the best ones overall.[31] The algorithm starts by showing a product. For now assume that one product is shown at a time. Fixing estimated search model parameters, and assuming one product per ranking page, we simulate consumers' utility from the shown product. If the utility is sufficiently high, then consumer clicks on a product and we record the payoff of one, and zero otherwise. The algorithm updates its posterior indices on products' performance $x^T\theta$. Next time a different consumer arrives, the algorithm shows either the same or some other product among eight available. Thus, every time a consumer gives feedback to the product, the algorithm updates and learns which product to show. Under sufficiently high number of simulations (consumers), the algorithm should learn which products are the best. We illustrate the results of a simple version of the algorithm with 2,000 iterations in Figure 16.

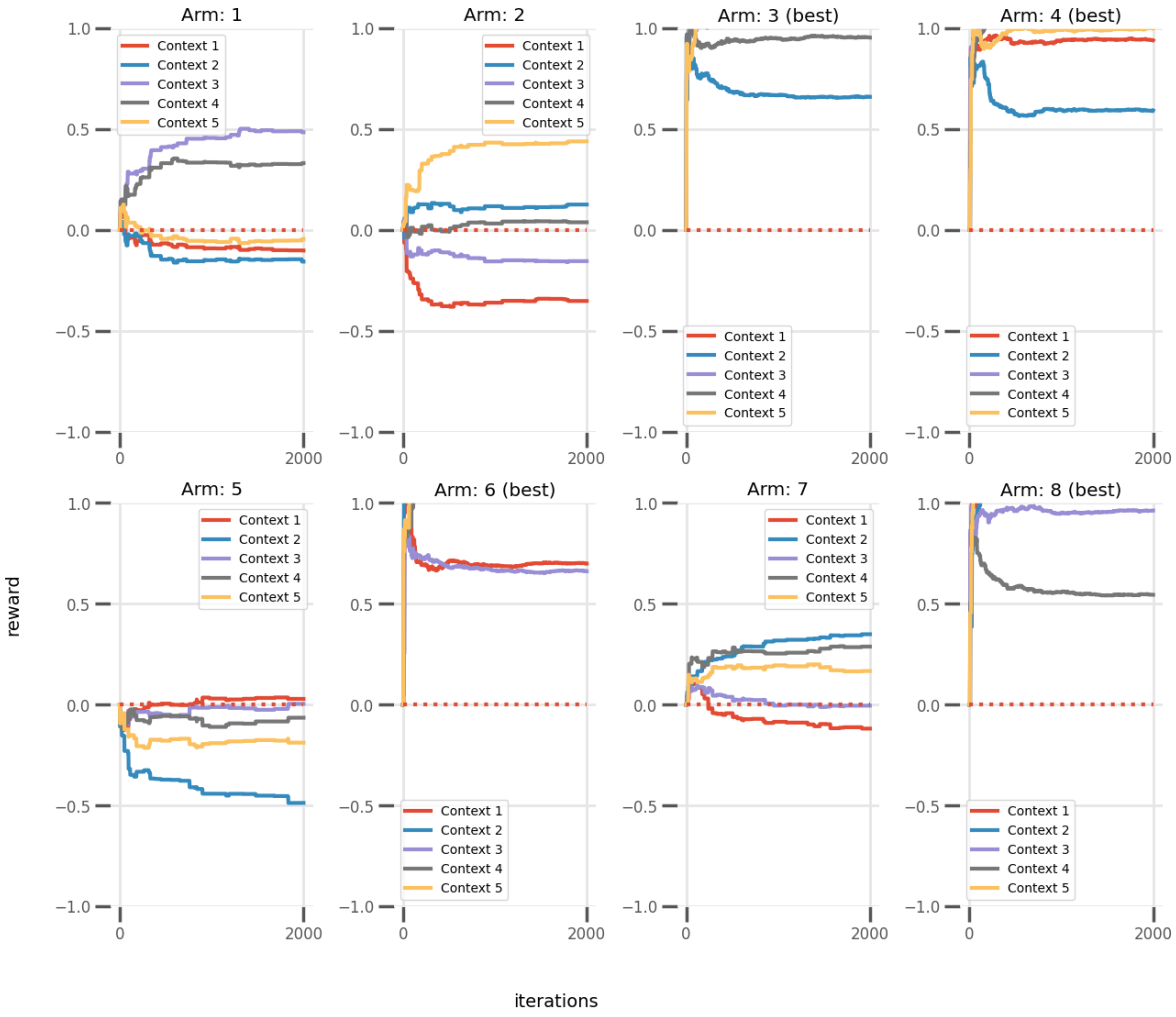
The x-axis of Figure 16 shows the iterations, y-axis shows the average reward, and different lines correspond to different contexts (geographic states). In the simulations, the algorithm learns quite quickly that arms 3, 4, 6, and 8 are the best arms and it starts playing them more often, which increases the overall rewards. Algorithm does not choose the remaining products as often because it learns that they are not liked (clicked) by the consumers. Overall, the more data the algorithm collects about the products' click-through-rates, the better it will learn which products to show next.

Recall that the optimal product is chosen via balancing exploitation and exploration an, i.e., showing better products and learning from uncertain products, following the optimal solution:

$$a_t = \underset{a \in A_t}{\arg\max}(x_{t,a}^T \hat{\theta}_a + \alpha\sqrt{x_{t,a}^T A_a^{-1} x_{t,a}}) \tag{17}$$

[31] We set the product attributes accordingly.

Figure 16 Illustration of the simple version of the algorithm

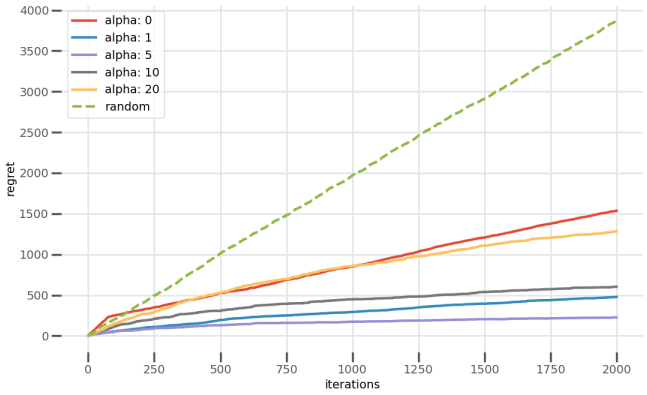


where α is an exploration parameter that plays an important role. If α is set too high, then the platform will show more uncertain payoff products, which may be good if the resulting payoff is high, but not otherwise.

We calculate the expected regrets from the simple algorithm with different exploration levels α and show the results in Figure 17. If $\alpha = 0$ then there is no exploration and regret would be high. However, under sufficiently high α, the regret drops considerably (lines below green dashed one). In our main counterfactuals, we run simulations using $\alpha = \{0, 5, 10, 20\}$.

Thus, the illustrative simulations show that the algorithm is able to find best products under sufficient number of iterations and reasonable exploration parameter α. In the main counterfactual, we follow the same procedure except that we have to generate the set of 72 products to form a ranking page for each consumer. Recall that to reduce the dimensionality of the problem, we show the same set of 72 product rankings within the same state. Our main outcomes of interest are the conversion probability, consumer welfare, platform's profit, and changes in the

Figure 17 Regret curves by exploration parameter α



environmental emissions. Table 10 shows the counterfactual results under different exploration rates α. Second column shows the conversion probability relative to the business-as-usual conversion rate. Under $\alpha = 0$ (no exploration) the bandit algorithm does not perform well and leads to 52% conversion (with benchmark business-as-usual conversion 100%).[32] As described above, low exploration parameter does not allow the algorithm to learn which products are best to show, and the algorithm keeps showing same set of products instead of exploring other products.

Table 10 Results from Bandit Simulations

Exploration parameter (1)	Conversion Probability relative to business-as-usual (%) (2)	Δ Consumer Welfare (3)	Platform Profit (4)
$\alpha = 0$	52.1	-\$12	-20%
$\alpha = 5$	67.9	-\$10	-14%
$\alpha = 10$	84.4	-\$0.5	-2%
$\alpha = 20$	99.7	\$0.8	+0.5%
Consumers	50,000		

[1] The baseline business-as-usual simulated welfare is \$10.

Next, we calculate the change in consumer welfare under bandit algorithm compared to the business as usual algorithm. To do that, we simulate consumer choices under bandit-generated rankings and calculate the maximum utility from purchasing, net of search costs. Next, we compare the welfare obtained under the bandit algorithm to the one that can be obtained under standard rankings. We repeat each simulation 50 times to integrate over the unobserved component of the utility. Third column in Table 10 shows that under $\alpha = 0$, the welfare decreases by \$12 compared to the business-as-usual ranking. Moreover, the platform's profit takes a negative hit given the decrease in conversion rates and decreases by 20%.

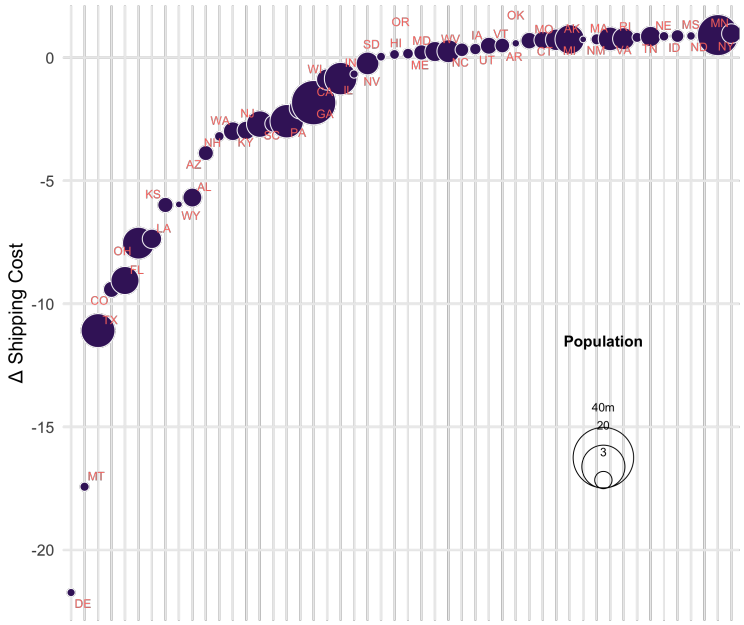
[32] Due to data sensitivity, we take the business-as-usual conversion as 100% and calculate the share of conversion that can be recovered using the algorithm.

As we increase the exploration parameter, the conversion probability starts to increase reaching 99.7% out of 100% under $\alpha = 20$. Even though conversion is still lower than the business-as-usual, consumer welfare is slightly higher (+\$0.8) and platform profit increases by 0.5%.

Thus, the proposed bandit algorithm can deliver almost as high conversion rate as usual and even increases platform's profit. The reason for the increase in platform's profit is that their shipping costs decrease under more environmentally-friendly rankings. Consumer welfare increases because we show products located closer to the consumers, and therefore, they have faster delivery, which consumers value.

The main remaining question is whether the algorithm is able to achieve environmental and shipping cost savings. Figure 18 shows the changes in the shipping cost by each state. The circles are proportional to the population in a state.

Figure 18 Counterfactual results



There is significant heterogeneity in the economies by each state. For instance, Delaware achieves the highest (scaled) decrease in shipping costs followed by Montana, Texas and Colorado. However, many states are either already at the lowest possible economies, such as New Jersey, which has many warehouses from which products can be shipped. Thus, for the focal platform and for other retail platforms, it is important to identify which regions may be good candidates with high potential of environmental and shipping cost decreases and potentially target those states with environmentally-friendly rankings.

Overall, the counterfactual results of both static and dynamic algorithms show that there is a potential to decrease environmental emissions without hurting conversion or profit metrics. Moreover, platforms' incentives are aligned with the environmental goals because lower CO2 emissions are proportional to lower shipping costs.

Limitations. One of the limitations of the analysis is that we measure user experience as the conversion probability. However, a more complete algorithm would also account for the probability of returning a product or the incidence rates. Technically, the incidents rates should be lower if the product travels less miles from the warehouse to the consumer. However, in this paper we ignore that effect because we do not know the counterfactual incidence rates. Similarly, we do not model product returns which is why we do not account for potential post-purchase product return. One could also be worried that in our calculations of the environmental emissions throughout the paper, we did not take into account that if consumers do not purchase then that would automatically decrease emissions because no item is being shipped. However, we assume that no purchase on the focal platform may result in a purchase on a competing platform, which will lead to potentially same level of emissions.

7. Conclusion

In this paper, we study the causal effect of fast delivery promises on environmental emissions, and propose alternative solutions for the platform to decrease environmental emissions. Despite regulatory concerns over the environmental impact of fast delivery promises, we find that removing fast delivery promises will lead to unintended consequences. We use the large-scale randomized experiment ran at Wayfair, where the platform randomly removed fast delivery promises for a large portion of traffic, to conduct a causal analysis of fast delivery promises. We find that removing fast delivery promises result in (i) substitution of demand towards products that were located 1% further away from the consumers, (ii) higher delivery wait times of 2.8%, (iii) higher prices and shipping costs. Moreover, removing fast delivery led to 1% higher emissions because of the miles traveled, as well as additional 2.5% emissions as a result of reduction in the economies of scale from centralized shipping. Thus, despite regulatory concerns, removing fast delivery promises not only does not decrease emissions, but it increases them, while also hurting consumer outcomes.

To help platforms reduce their environmental emissions, we propose a static and dynamic multi-arm bandit-based ranking algorithms that balance two goals: (i) maximizing purchase probability and (ii) minimizing environmental emissions. To evaluate the algorithms, we estimate a sequential search model to recover consumer search costs and preference parameters. We then simulate how consumers respond to alternative set of rankings, fixing estimated model

parameters. We show that the algorithms can maintain same level of purchase probability as in the business-as-usual case, and can decrease environmental emissions and platform shipping costs.

Overall, the results in the paper suggest that one way of decreasing environmental emissions is to nudge consumers towards more environmentally-friendly products through product rankings on the retail platforms. An interesting future direction in this area could be experimenting with different messaging on the product ranking pages, that would make consumers aware of the environmental goals behind product rankings.

References

Auer, Peter, Nicolo Cesa-Bianchi, and Paul Fischer, "Finite-time analysis of the multiarmed bandit problem," *Machine learning*, 2002, *47*, 235–256. (Cited on page(s) 4)

Bollinger, Bryan, Randi Kronthal-Sacco, and Levin Zhu, "Sustainable Product Demand and Price Elasticity," 2023. (Cited on page(s) 3)

Brecko, Kristina and Yewon Kim, "Voluntary Provision of Sustainability Claims: Evidence from Consumer Packaged Goods," Technical Report, Working paper 2023. (Cited on page(s) 3)

Chen, Yuxin and Song Yao, "Sequential search with refinement: Model and application with click-stream data," *Management Science*, 2017, *63* (12), 4345–4365. (Cited on page(s) 25)

Cristea, Anca, David Hummels, Laura Puzzello, and Misak Avetisyan, "Trade and the greenhouse gas emissions from international freight transport," *Journal of environmental economics and management*, 2013, *65* (1), 153–173. (Cited on page(s) 3)

Duan, Huabo, Guanghan Song, Shen Qu, Xiaobin Dong, and Ming Xu, "Post-consumer packaging waste from express delivery in China," *Resources, Conservation and Recycling*, 2019, *144*, 137–143. (Cited on page(s) 3)

Hodgson, Charles and Gregory Lewis, "You Can Lead a Horse to Water: Spatial Learning and Path Dependence in Consumer Search," 2022. (Cited on page(s) 25)

Juan, Angel A, Javier Faulin, Elena Pérez-Bernabeu, and Nicolas Jozefowiez, "Horizontal cooperation in vehicle routing problems with backhauling and environmental criteria," *Procedia-Social and Behavioral Sciences*, 2014, *111*, 1133–1141. (Cited on page(s) 3)

Kang, Peng, Guanghan Song, Ming Xu, Travis R Miller, Haikun Wang, Hui Zhang, Gang Liu, Ya Zhou, Junshu Ren, Ruoyu Zhong et al., "Low-carbon pathways for the booming express delivery sector in China," *Nature Communications*, 2021, *12* (1), 450. (Cited on page(s) 3)

Kellogg, Ryan and Mar Reguant, "Energy and environmental markets, industrial organization, and regulation," in "Handbook of Industrial Organization," Vol. 5, Elsevier, 2021, pp. 615–742. (Cited on page(s) 3)

Kim, Jun B, Paulo Albuquerque, and Bart J Bronnenberg, "Online demand under limited consumer search," *Marketing science*, 2010, *29* (6), 1001–1023. (Cited on page(s) 24, 26)

Li, Lihong, Wei Chu, John Langford, and Robert E Schapire, "A contextual-bandit approach to personalized news article recommendation," in "Proceedings of the 19th international conference on World wide web" 2010, pp. 661–670. (Cited on page(s) 4, 22, 23)

Loon, Patricia Van, Lieven Deketele, Joost Dewaele, Alan McKinnon, and Christine Rutherford, "A comparative analysis of carbon emissions from online retailing of fast moving consumer goods," *Journal of Cleaner Production*, 2015, *106*, 478–486. (Cited on page(s) 3)

Muñoz-Villamizar, Andrés, Josué C Velázquez-Martínez, Perla Haro, Ana Ferrer, and Roger Mariño, "The environmental impact of fast shipping ecommerce in inbound logistics operations: A case study in Mexico," *Journal of Cleaner Production*, 2021, *283*, 125400. (Cited on page(s) 3, 8)

Sanders, Robert Evan, "Dynamic Pricing and Organic Waste Bans: A Study of Grocery Retailers' Incentives to Reduce Food Waste," *Available at SSRN 2994426*, 2020. (Cited on page(s) 3)

Santos, Babur De Los, Ali Hortaçsu, and Matthijs R Wildenbeest, "Search with learning for differentiated products: Evidence from e-commerce," *Journal of Business & Economic Statistics*, 2017, *35* (4), 626–641. (Cited on page(s) 25)

Tran, Nguyen Khoi and Jasmine Siu Lee Lam, "Effects of container ship speed on CO2 emission, cargo lead time and supply chain costs," *Research in Transportation Business & Management*, 2022, *43*, 100723. (Cited on page(s) 3)

Ursu, Raluca M, "The power of rankings: Quantifying the effect of rankings on online consumer search and purchase decisions," *Marketing Science*, 2018, *37* (4), 530–552. (Cited on page(s) 2, 24, 25, 27)

Weitzman, Martin, *Optimal search for the best alternative*, Vol. 78, Econometrica, 1979. (Cited on page(s) 2, 24, 25)

A Additional Descriptives

Table 11 Randomization check (see p. 12)

	Control (1)	Treatment (2)	Difference (3)	p-value (4)
Historical Purchases ($)	277.82	261.53	16.29*** (10.98)	0.00
Historical Quantity Bought	2.32	2.20	0.12*** (11.39)	0.00
Estimated networth	426,454.69	425,257.17	1,197.52*** (4.27)	0.00
Estimated income	69,545.70	69,315.94	229.76*** (8.77)	0.00
Age	26.82	26.83	-0.01** (-2.94)	0.00
Home Value	175,121.26	174,672.35	448.92*** (4.33)	0.00
Prices searched	73.73	73.47	0.26*** (4.41)	0.00
Gender dummy (0,1)	0.93	0.93	-0.01*** (-29.40)	0.00

Notes: This table shows randomization checks.

Figure A.19 Consumer search process (see p. 5)

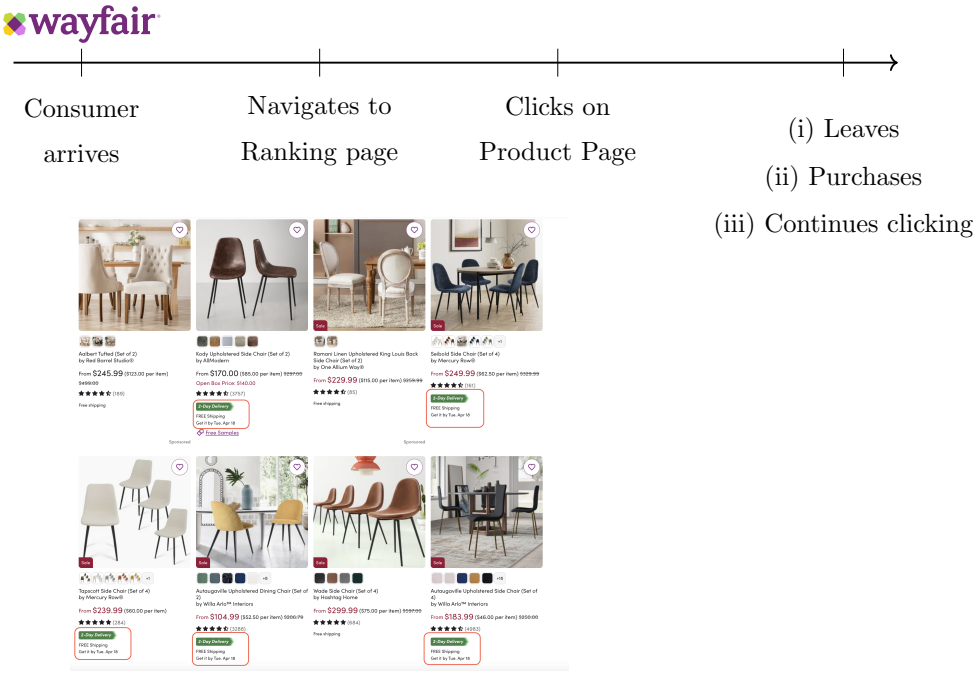


Figure A.20 Percent of good quality air days by US states

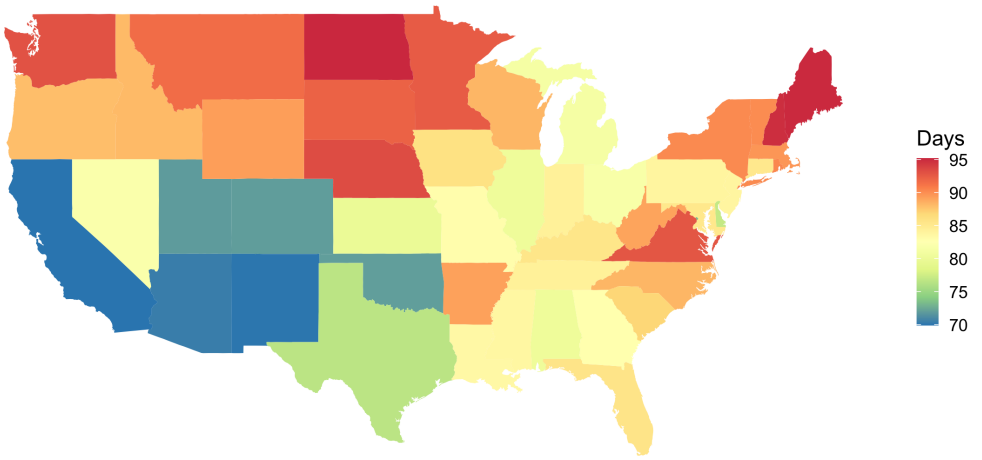
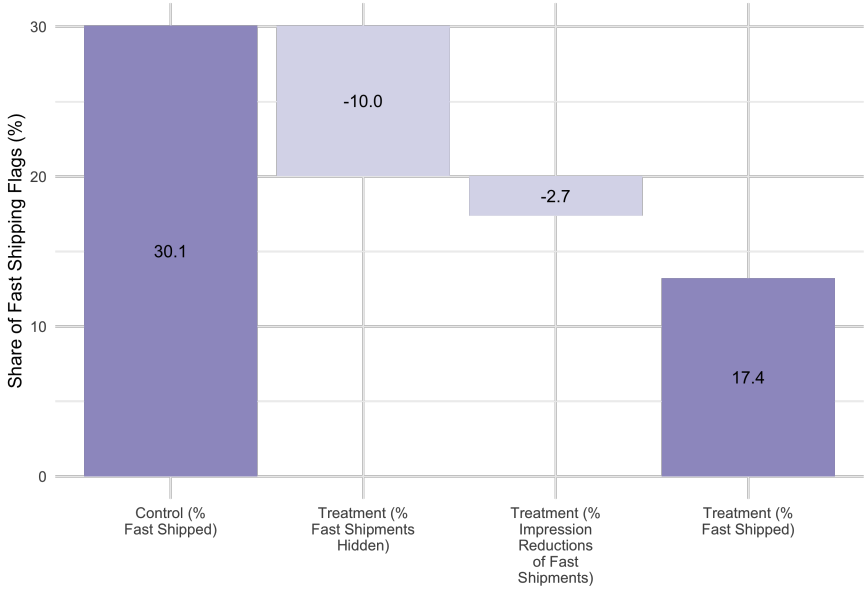


Figure A.21 shows that the experimental impact was very similar in the dining chair category with 42% reduction in fast delivery promises. This validates that the dining chairs is representative of the experiment.

Figure A.21 Share of fast shipped products between control and treatment groups



B Centralized distribution centers

Table 12 Effect of centralized distribution on delivery outcomes

	(1) miles traveled	(2) shipping costs	(3) transit days	(4) delay days	(5) miles traveled
Distribution Center	-425.431***	-0.653***	-1.732***	-0.773***	-282.144***
	(0.745)	(0.088)	(0.293)	(0.293)	(1.195)
Fast Badge					-62.044***
					(0.496)
Distribution Center× Fast Badge					-455.256***
					(2.048)
Intercept	1133.858***	19.226***	8.656***	1.578***	1161.789***
	(0.157)	(0.018)	(0.064)	(0.064)	(0.254)
Observations	21,565,998	22,602,359	20,143,114	20,132,571	12,888,542

Notes. This table shows that the centralized distribution centers decrease miles traveled by a product by 425 miles or by 37%, they also decrease shipping costs, transit days and delay days. The effect is even bigger for the fast-shipped products. OLS regression with robust standard errors. Statistical significance: $^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$.

Table 13 Effect of centralized distribution on weighted-miles

	(1) miles traveled × weight	(2) miles traveled × weight
Distribution Center	-31,250.290***	-16,938.031***
	(89.859)	(188.522)
Fast Badge		-4,111.247***
		(78.276)
Distribution Center× Fast Badge		-34,390.752***
		(323.136)
Intercept	80,529.476***	77,084.201***
	(18.992)	(40.103)
Observations	21,565,998	12,889,033

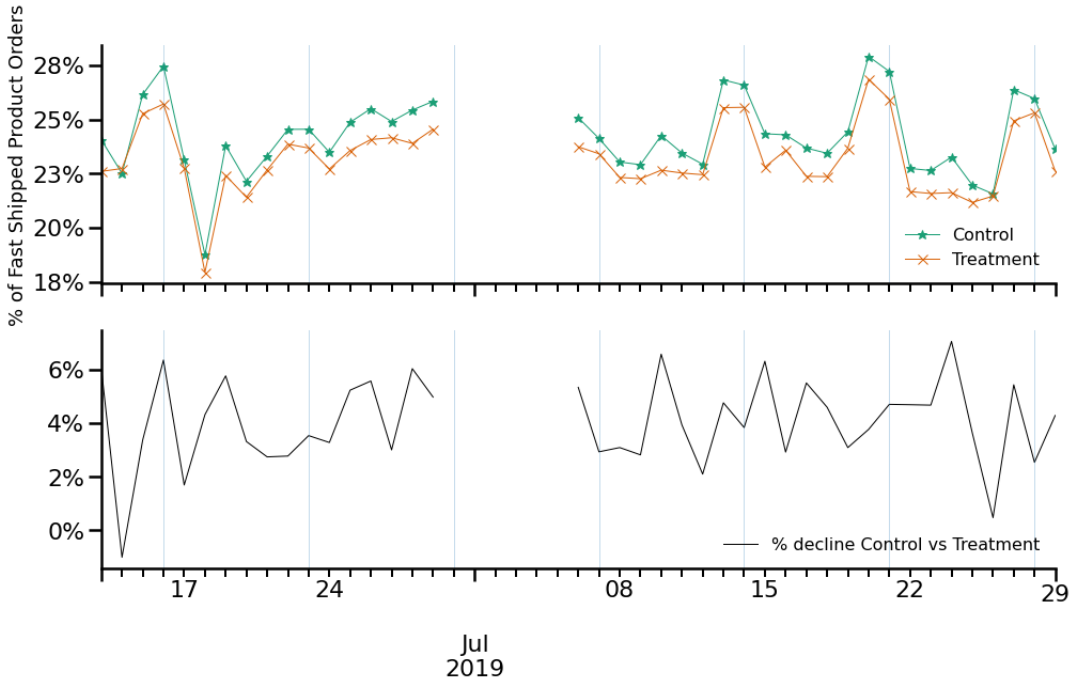
Notes. This table shows that the centralized distribution centers decrease miles-ton by 38.8% or 31,250 miles-ton. OLS regression with robust standard errors. Statistical significance: $^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$.

Table 14 Effect of centralized distribution by urban locations

	(1) miles traveled	(2) shipping costs
Distribution Center	-71.287***	-102.294***
	(1.506)	(0.440)
Metropolitan Area	-0.220	
	(1.223)	
Distribution Center× Metropolitan Area	-33.190***	
	(1.571)	
Adjacent to Metropolitan Area		-28.262***
		(1.529)
Distribution Center× Adjacent to Metropolitan Area		11.527***
		(1.958)
Intercept	896.953***	898.160***
	(1.173)	(0.341)
Observations	10,772,350	10,772,350

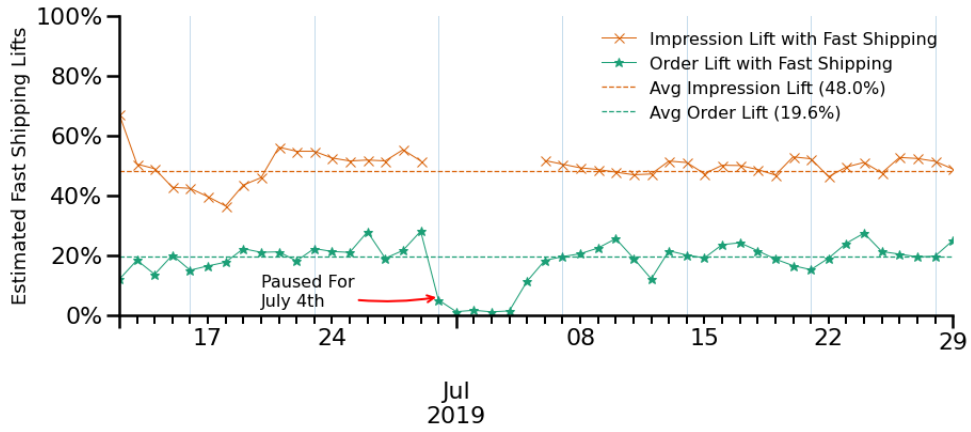
Notes. This table shows that the centralized distribution centers decrease the miles traveled by a package in urban (metropolitan) area by 33 miles or by 3.7% more than in non-metropolitan area. The effect does not persist in the non-urban or adjacent to metropolitan areas. OLS regression with robust standard errors. Statistical significance: $^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$.

Figure B.22 % of Fast Shipped Product Orders



Notes. The figure shows that the fast-shipped orders decreased in the treatment (reduced fast shipping) condition.

Figure B.23 Estimated Shipping Lifts



Notes. The figure shows that fast shipping badges increase product impressions on the ranking pages, and they lead to 17-18% order lift (mirror results to Table 6).