Enhancing Local Fulfillment in Retail: A Structural Model of Fulfilling Demand Sensitive to Waiting

Dayton Steele
Kenan Flagler Business School, University of North Carolina at Chapel Hill

Saravanan Kesavan
Kenan Flagler Business School, University of North Carolina at Chapel Hill

Abstract

JD.com leverages a regional network of distribution centers (DCs) to fulfill online orders, consisting of regional DCs and front DCs. Front DCs improve delivery times for their local customers due to closer proximity. However, front DCs only hold the required inventory to fulfill locally for 35% of orders, requiring them to leverage backup fulfillment from regional DCs. Backup fulfillment impacts the customer experience, with a 1.1 day increase in average promised delivery time. We show that customers are sensitive to the promised time and longer promise times decrease sales. Thus, JD.com faces a central trade-off: stock the local DC to maximize sales but increase replenishment costs; or leverage the network to lower operating costs but sacrifice sales due to increased promise times. We build a structural model where the decision maker acts rationally according to the newsvendor model in balancing the trade-off of fulfilling demand locally or through backup fulfillment. We estimate relative costs for local fulfillment, supporting the empirical evidence that front DCs fulfill less orders locally than regional DCs. We use our model to explore counterfactual scenarios at the front DCs, given their reliance on backup fulfillment. We find that JD.com's current utilization of front DCs drives 16% in decreased median promised time for local demand, resulting in 1.8% increased revenue for local demand. If front DCs fulfilled all of their local demand, median promised time would decrease by an additional 33% leading to an additional 4.3% increase in revenue. However, doing so would require substantial investment in front DCs to lower local fulfillment costs. We explore the sensitivity of revenue gains to short-term investments that reduce local fulfillment costs.

1 Introduction

JD.com distinguishes itself in the Chinese eCommerce market with its superior logistics. JD.com's self-operated nationwide logistics network provides a key competitive advantage in its ability to offer 90% same-or-next-day delivery as a standard service, while still maintaining low distribution costs. As stated by Sidney Huang, CFO of JD.com, "Mainly, our quick delivery is a result of our warehouse network, which means the products can be extremely close to our customers" (Zhu and Sun 2019). One key component from JD.com's logistics network is the setup of distribution centers in order to minimize the number of times goods move around, typically reduced from four to five movements in traditional logistics, to one or two movements maximum (see Zhu and Sun (2019) for more details). JD.com leverages a two-tier[1] distribution center network, displayed in Figure 1, consisting of regional DCs and front DCs (Ma et al. 2018). Regional DCs have large storage capacity but are fewer in number; front DCs can reach customers in surrounding areas directly but have less storage capacity. The closest DC attempts to fulfill demand directly depending on the proximity of the customer. When the closest DC does not have the required inventory to meet its local demand, it may use backup fulfillment by requesting assistance from another DC. Since backup fulfillment requires shipping from a DC further from the customer, the promised delivery time increases. Thus, JD.com faces a central problem: how to best fulfill local demand in each DC in order to minimize delivery speed to maximize sales, but balance the costs of local fulfillment compared to backup fulfillment. Local fulfillment costs may include logistics costs of frequent replenishment or administrative warehouse costs of holding inventory, whereas backup fulfillment costs may include increased shipping costs. Furthermore, demand is realized after the point of inventory replenishment, so JD.com makes its fulfillment decisions with uncertain demand for each product. Perakis et al. (2020) propose an algorithm that uses the classical Operations Management (OM) newsvendor model[2] to achieve optimal allocation across different DCs in a similar online fulfillment setting. Their algorithm requires inputs of "underage" costs for replenishing too little inventory and "overage" costs for replenishing too much inventory, at the DC-product level. However, these costs are unobserved so it is not clear how to estimate these inputs. We similarly model local fulfillment decisions under the newsvendor framework and provide a methodology to obtain these cost inputs. Using our model, we evaluate the gains to JD.com from enhancing local fulfillment through front DCs to improve delivery times to improve sales.

The data provided by JD.com in the 2020 MSOM data competition allows us to more closely examine this problem. In particular, the data provides transactional data of customer orders mark-

[1]In practice JD.com's regional network is more complicated but operations reflect the premise of fewer touch-points in distribution. For example, Zhu and Sun (2019) points out a three-tier distribution network. We focus on the simpler two-tier network for exposition reasons and because it aligns with the data given.

[2]For readers unfamiliar with the newsvendor framework, Porteus (2002) is an excellent resource.

Figure 1: JD's Two-Level Distribution Network

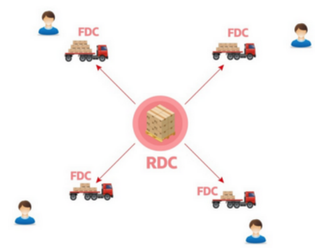


Figure copied from Ma et al. (2018)

ing the closest DC for fulfillment ("dc_des") and the actual DC that fulfilled the order ("dc_ori"). We can see that regional DCs have very high service levels compared to front DCs, fullfilling orders locally for 93% of orders compared to 35% of orders at front DCs.[3] Clearly front DCs burden larger costs in fulfilling locally than regional DCs, but these costs are not observable in the data. In addition to difficulty in examining costs in the local fulfillment decision, quantifying the impacts to revenue from local fulfillment decisions is not straightforward. Local fulfillment impacts demand indirectly through promised delivery time, whereas in a traditional retail setting local fulfillment may impact demand directly through product availability. We capture the local fulfillment decision that incorporates indirect effects to demand through promised delivery time, balancing unobserved costs of local fulfillment.

We build a structural model[4] to capture the equilibrium state of the system for JD.com, and to examine changes to the system in leveraging front DCs for local fulfillment. We assume that the manager at each DC behaves optimally in choosing its local fulfillment. The current choice of daily local fulfillment optimally balances the costs of fulfilling locally and improving sales through better delivery times. We model the manager's optimal decision under the classical OM framework of the newsvendor model. The manager incurs "underage" costs for replenishing too little inventory and "overage" costs for replenishing too much inventory. We demonstrate that customers are sensitive to the delivery time, which is impacted by the local fulfillment decision, which is impacted by the relative cost of local fulfillment. Perakis et al. (2020) make similar modeling assumptions to apply

[3]Throughout this paper we use statistics from our working sample, which drops data with missing attributes. Across all data provided by JD.com, the fulfillment statistics are comparable, with local fulfillment for regional and front DCs at 96% and 33% respectively. More details are given in the Data section.

[4]For readers unfamiliar with structural estimation, the goal of a structural model is to specify the decision-making process of the agent, assume the agent behaves optimally under this framework, and then use the data to estimate the primitives of the model that drive behavior. Because the model captures "primitives" of behavior, we can make changes to the system that are not endogenous to the outcomes of the system. Reiss and Wolak (2007) is an excellent introduction to this approach.

the newsvendor model to online retailing in a two-tier fulfillment network. Unlike Perakis et al. (2020) who use an analytical approach to derive normative results, our goal is to deliver managerial insights by revealing unobserved costs. In order to achieve this, we apply a structural modeling approach, similar to Olivares et al. (2008) who use structural modeling to reveal unobserved costs of reserving operating room time. Unlike in Olivares et al. (2008), demand is impacted by the service level, indirectly through promised delivery time. We extend Olivares et al. (2008) by endogenizing demand through promised delivery time, which is impacted by service level.

Structural modeling is required over reduced-form estimation, or simply estimating the statistical relationships among data of interest, in addressing our research question of quantifying the benefits of front DCs and how to improve their utilization at JD.com. As pointed out in Reiss and Wolak (2007), structural estimation should be used to estimate unobserved behavioral parameters that cannot be inferred from nonexperimental data. In particular, the costs limiting front DCs from higher fulfillment rates are not observable in the data, and our behavioral assumptions on the optimal decision-making for local DC managers allows us to specify and estimate the cost parameters across DCs. These unobserved costs provide the mechanism for how policy changes can realistically occur. For example, while identifying the value of delivery times to improved revenue is interesting by itself, we believe the greater interest for JD.com lies in how to best leverage its resources to improve delivery times to then improve revenue. By specifying the detailed relationship between local fulfillment costs and the DC's optimal decision in forming expectations on delivery time improvements and revenue, we can support realistic policy improvements with theory. In other words, we propose that delivery times are endogenous to the local DC fulfilling the order, and cannot be directly improved without improving in-stock probability, requiring costly enhancements to the local DC. These reasons support the general point that structural models should be used for counterfactual experiments, also discussed in Reiss and Wolak (2007).

Our analysis is useful to JD.com for the following reasons. We characterize a supply-side difficulty to improving promised delivery times. Our estimated cost parameters can be used as inputs in conjunction with algorithmic approaches such as Perakis et al. (2020) to improve inventory allocation. We explore the value of leveraging front DCs to improve delivery times by closer proximity to certain customers than regional DCs. We capture the total opportunity from leveraging front DCs for fulfillment, by satisfying all local demand at the front DC instead of all local demand through regional DCs, to be a 43% reduction in median promise time resulting in a 6.1% increase in front DC revenue. We find that JD.com's current utilization of front DCs drives 16% in decreased median promised time for local demand, resulting in 1.8% increased revenue for local demand. If front DCs fulfilled all of their local demand, median promised time would decrease by an additional 33% leading to an additional 4.3% increase in revenue. However, doing so would require substantial investment in front DCs to lower local fulfillment costs. We explore the revenue gains from less substantial short-term improvements to local fulfillment costs. We also examine which front DCs could benefit most from improvements to local fulfillment costs, identifying the highest potential increases in revenue after taking into account consumers' sensitivity to delivery time. We show that those DCs with the most upside for short-term investment through practical short-term cost reductions are not the same as those DCs with the most upside for long-term investment through total remaining opportunity in revenue.

We leverage the data presented by JD.com to address our research question: what is the additional value in enhancing local fulfillment in online retail through closer distribution centers, and how can the retailer make enhancements to improve their utilization of closer distribution centers?

2 Related Work

Recent OM empirical studies have demonstrated that consumers respond positively to reduced delivery time. Using a quasi-experiment in an omnichannel retail environment, Fisher et al. (2019) show that on average sales increase by 1.45% per business-day reduction in delivery time. In a quasi-experiment at Alibaba, Cui et al. (2019) show that the removal of high-quality delivery partner SF Express negatively impacted sales by 14.56%. We similarly observe customer sensitivity to delivery time in JD.com's data. Our approach is to leverage a JD.com manager's expectation of the impact on delivery time on sales for informing the local fulfillment decision.

Optimal inventory allocation has been widely studied in the OM community. The general approach is to solve an optimization problem, such as a linear program, that accounts for costs of allocation decisions while incorporating constraints on demand and capacity. The allocation problem has gained recent attention in specific application to the online retailing environment. For example, Chen and Graves (2020) formulate a mixed-integer program that solves an inventory placement problem for an online retailer in choosing which fulfillment centers to place items. Perakis et al. (2020) use Lagrangian duality to propose an algorithm for solving a newsvendor model balancing overage and underage costs for distribution centers in a two-tier fulfillment network for an online retailer. However, these approaches generally require knowledge of the cost parameters in solving the optimization problem, and can be very costly to solve across many products and fulfillment centers. Instead of solving for the optimal allocation across the entire network, we instead focus on each DC's fulfillment decisions for each SKU on a given day. We model each DC manager as solving a newsvendor problem for each SKU on each day. Instead of taking the cost parameters as given, we estimate them using a structural model.

Structural modeling is recent to the OM community. Terwiesch et al. (2020) provide a thorough review of the increased emphasis of empirical work in OM, pointing out examples of those using structural modeling. Our approach is most similar to Olivares et al. (2008) who model hospital decisions in allocating operating room time, balancing the cost of reserving too much or too little operating room capacity. In their context, demand is not a function of the service level. However, for JD.com, the inventory allocated to fulfill demand locally also impacts sales through its impact on delivery time. Other structural papers present contexts with some rough similarities to that of JD.com. Akşin et al. (2013) model caller sensitivity to delay in call centers, similar to customer sensitivity to delivery times at JD.com. Allon et al. (2011) model fast-food restaurant to show that customers have a high cost to waiting for service. Both papers suggest that the firm should incorporate customer reaction to waiting times into operational decisions. Musalem et al. (2010) estimate the effect of lost sales of stockouts, similar to the negative effect on sales of increased delivery times from stockouts in a local DC. In this sense, the effect of stockouts for JD.com is different: increased delivery times mitigate the full effect of a stockout, and remove the concern of studying censored demand data. While these structural papers provide insights that could translate to JD.com, none of them model a distribution center manager's daily local fulfillment decision in a regional fulfillment network.

To our knowledge, our work is the first to empirically examine the managerial decision of fulfilling demand locally or leveraging backup fulfillment as it impacts customer delivery time in a two-tier distribution network. We quantify the benefits of front DCs and examine policy changes to capture these benefits. We leverage our model to examine the potential impacts of using front DCs to improve promise times, and thus improve revenue, relative to shipping directly from regional DCs.

3 Data

We leverage data provided by JD.com in the 2020 MSOM data competition. JD.com provides access to seven linked data tables: clicks, deliveries, inventory, network, orders, SKUs, and users. As noted in the prompt for the competition, many observations have attributes missing, mostly on third-party SKUs for which JD.com has less data. In particular, these attributes are commonly missing in the data set: promise time on the order (37% of unique orders), delivery information for the order (40% of unique orders), customer attributes (33% of unique orders), click data (28% of unique orders), SKU quality attributes (21% of unique orders), and final prices (14% of unique orders). In forming our model we use all of these data attributes, so we filter the data set for only those observations with data on each attribute. The original data set has 486,928 unique customer orders with 62% of SKUs ordered as Type 1. Our working data set has 207,549 unique customer orders with 87% of SKUs ordered as Type 1.

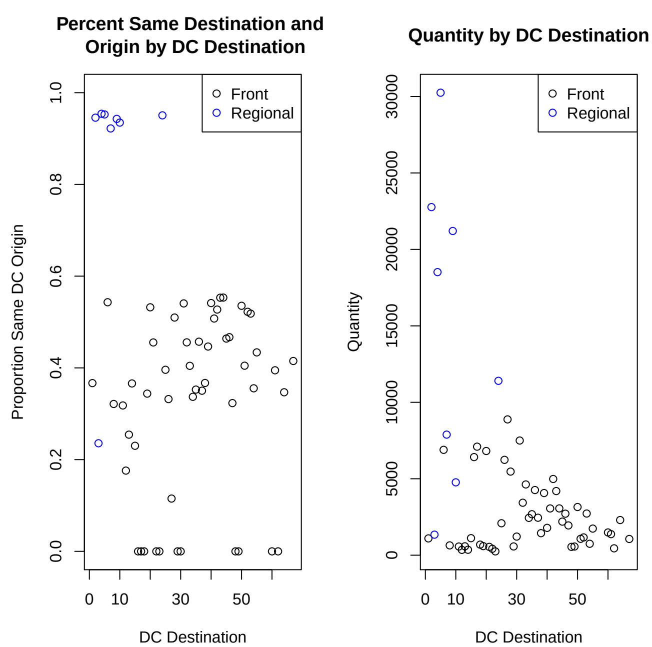
Table 1 presents summary statistics by whether the DC is the local DC or not on a customer order. We can see that in our working sample, 63% of orders are filled locally. Type 1 SKUs, or SKUs managed through JD.com, are more likely to be filled locally, as shown from the difference of 91% and 79% in the table. We see that the average final price for local fulfillment is higher, due to a different mixture of SKUs. We can see that the average promise time is 1.1 days faster when delivered locally, and both the average promise and average delivery time take about half the time when fulfilled locally. Both the average purchase power and percentage of PLUS customers are about the same whether filled locally or regionally. We found this to be true across other customer attributes as well. Table 1 gives high level evidence that promise time is impacted by JD.com's local fulfillment decisions.

Table 1: Summary Statistics by Local Fulfillment

	Local	Non-Local	All
Total Orders	130,934	76,615	207,549
Total Sales	156,972	95,364	252,336
Average Price (in RMB)	99.73	83.16	93.62
Average Promise	1.41	2.51	1.82
Average Delivery	0.88	1.56	1.13
Percent Type 1	0.91	0.79	0.87
Average Purchase Power	2.25	2.30	2.27
Percent PLUS	0.25	0.22	0.24

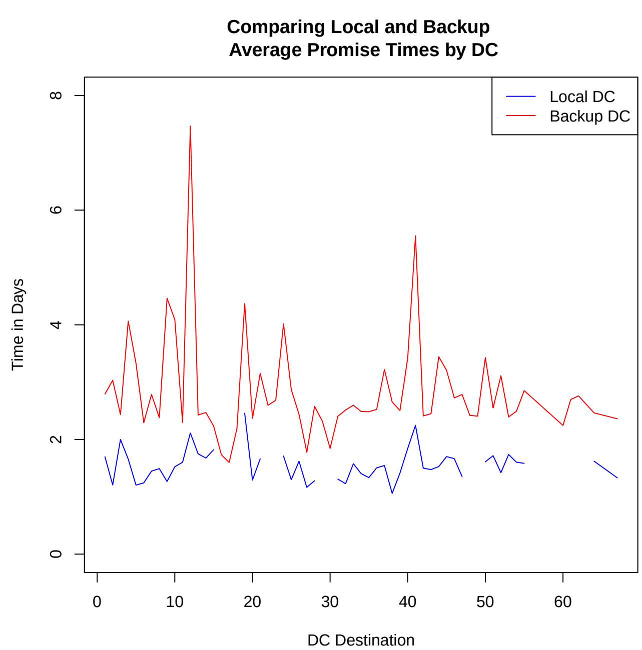
The left panel of Figure 2 shows the proportion of local DC fulfillment by DC. In blue are the cases when the regional DC serves to fulfill demand locally. Regional DCs are identified when they have an ID from the following: 2, 3, 4, 5, 7, 9, 10, 24. We can see that when the regional DC is closest to the customer, it almost always chooses to fulfill demand locally at close to 100% local fulfillment. On the other hand, the highest local fulfillment rates for front DCs is 55%. Front DCs heavily leverage regional DCs for assistance in fulfillment, with heterogeneity in levels of fulfillment. In our sample, eleven DCs leverage back-up fulfillment for all of their local orders; in the original sample these DCs leverage back-up fulfillment for about 85% of their local orders. The right panel shows the total quantity from which the DC is the local DC. We can see that both regional and front DCs show substantially more heterogeneity in total sales compared to their fulfillment rates. For example, some regional DCs represent as much local demand as front DCs but have much higher local fulfillment rates. This provides early evidence that the DC fulfillment decisions are not entirely demand driven – cost side factors also influence each DC's ability to fulfill demand locally.

Figure 2: Proportion of local DC fulfillment by DC, and total market share of sales by DC.



There is also heterogeneity in the promise time given the local DC, whether fulfilling the order locally or fulfilling the order through back-up fulfillment. Figure 3 plots the average promise time for a DC when it fulfills its demand as the local DC and the average promise time when it fulfills its demand from a backup DC. As we would expect, we see that the average promise time when fulfilling locally is always lower than the promise time when fulfilling through a backup DC. The average promise time when fulfilling locally is generally less than two days, but there are some slight differences across DCs. This suggests that the placement of DCs roughly aligns with equalizing the local promise time across DCs. Note that there are gaps in the promise time for local DCs. This is because some DCs in the data never fulfill orders locally, as shown in Figure 2, so that we do not have a measure for their local promise time. On the other hand, all DCs leverage backup fulfillment at some point allowing us to measure it across all DCs. The backup DC fulfillment has considerably more variance across DCs, with some DCs expecting close to eight days in back up fulfillment, whereas others have backup fulfillment closer to the local DC fulfillment. Since the data competition does not give the locations of the DCs, we can think of the average promise time from backup fulfillment as a proxy for shipping distance from the backup DC. Assuming that shipping distance incurs a cost, as discussed in Perakis et al. (2020), we may expect that those DCs with lower average promised times from backup fulfillment, relative to average promised times from local fulfillment, have a lower cost for backup fulfillment. In the newsvendor framework this is represented by the underage cost, implying that the relative costs of local fulfillment would increase because the opportunity cost for backup fulfillment is low. Thus, Figure 3 presents additional evidence for differing costs across DCs in local fulfillment decisions.

Figure 3: Comparison of Promise Times When Fulfilled Locally or from Backup



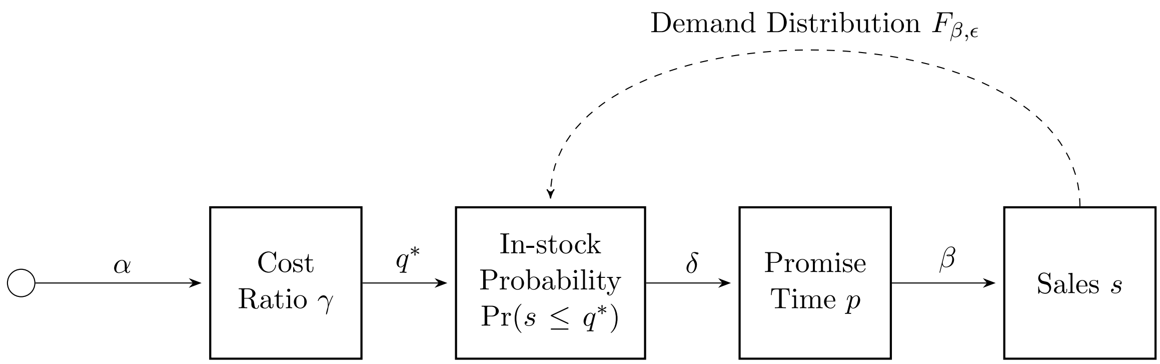
4 Model

We model a straightforward decision-making process for each manager at a given local DC for JD.com: how many units of sales do I fulfill locally per SKU? Given the regional network, we assume that all sales are met. We can think of shipping from another DC as analogous to a back-order. The manager must set the order quantity at the start of the day prior to the realization of sales. In practice, this order quantity could be some combination of inventory remaining the night before and new units shipped on a daily truck from regional DCs. In JD.com's data we do not observe the actual units in inventory so we cannot know what portion is shipped that day or held overnight. To account for this data issue, we assume that all local fulfillment in a given day is due to daily replenishment, abstracting away from the dynamics of inventory held into the next day. Perakis et al. (2020) similarly abstract away from multi-period dynamics. The manager forms rational expectations on sales and promised delivery time based on the realizations in the data. Expectations of sales for each SKU in the day are a function of the expected promised delivery time. The expected promised delivery time is a function of the in-stock probability, or expected service level, set by the manager's chosen order quantity. As the in-stock probability increases, the proportion of inventory fulfilled locally increases so that the expected promised delivery time decreases. Once the order quantity is chosen based on the expected demand from the expected promised delivery time, demand is realized. Figure 4 presents a diagram for the mechanics of our model, to be defined throughout the remainder of this section. We can see that the manager's optimal decision q^* takes into account both its relative costs from the cost ratio and its decision in relation to demand, which is intimately connected through the promise time. We will refer to those items that drive underlying behavior as *primitives*, e.g. α, δ, β, and those items that are realizations of unobservables and behavior of the agents as *outcomes*, e.g. γ, q^*, p, s.

Let q_{ijt} be the number of units ordered to fulfill sales s_{ijt} with daily promise time p_{ijt} for local distribution center $i \in C$, for SKU $j \in J$, in day $t \in T$. Moving forward, we will refer to vectors with undercase letters and matrices with uppercase letters. Thus, a_{ijt} is a scalar, whereas A_{ijt} is a row vector. We model s_{ijt} with a log-linear relationship to observed drivers of demand X_{ijt}, which can be written as

$$\begin{aligned} \log s_{ijt} &= X_{ijt}\beta + \epsilon_{ijt} && (1) \\ &= X_{jt}^{clicks}\beta_{clicks} + X_j\beta_j + p_{ijt}\beta_{prom} + price_{ijt}\beta_{price} + && (2) \\ &\quad X_i^{FE}\beta_i + X_t^{FE}\beta_t + \epsilon_{ijt} \end{aligned}$$

Figure 4: Diagram of Model for Local Fulfillment Decision



where β is a column vector of parameters and ϵ_{ijt} is a scalar distributed with mean zero and standard deviation of σ_ϵ. We assume ϵ_{ijt} is a factor that shifts demand which is observed by the customers and manager, ex-post the quantity decision is made, but unobserved to the researcher. Our demand drivers include the volume and characteristics of interested customers contained in the clicks table for SKU j on day t, characteristics of SKU j captured in the SKU table (including whether the SKU is owned by JD.com labeled with "Type 1" or through a third party seller), a measure of promised delivery time for the given observation, the average observed final price for the observation which is assumed to be set exogenously, and fixed effects for customers with local DC i and day t. To improve model fit, we could incorporate fixed effects per SKU. However, the number of SKUs with sales is approximately 1000 so this approach is very costly. Furthermore, this approach could lead to overfitting and removes our ability to interpret the effects of SKU characteristics. Instead, we attempt to capture the characteristics of the SKUs from the SKU table.

For our measure of promised delivery time we examine the median promised time for the observation. We choose the median because it better reflects what the standard customer should expect, and better fits our data both in estimation and in predicting the equilibrium. Let $InStockProb$ be the in-stock probability for the observation, determined by the manager's fulfillment decision. We model the median promise time p_{ijt} with a log-linear relationship to observed drivers of delivery V_{ijt}, which can be written as

$$p_{ijt} = V_{ijt}\delta + \omega_{ijt} \tag{3}$$

$$= V_j\delta_j + InStockProb_{ijt}\beta_{InStockProb} + V_i^{FE}\delta_i + V_t^{FE}\delta_t + \omega_{ijt} \tag{4}$$

where δ is a column vector of parameters and ω_{ijt} is a scalar distributed with mean zero and standard deviation of σ_ω. We assume ω_{ijt} is a factor that impacts the promise time observed to the manager but unobserved to the researcher. Our promise time drivers include SKU-specific characteristics that may impact the time to ship the product, in-stock probability which impacts the chance that the SKU will be non-locally fulfilled with a longer delivery time, and fixed effects for the local DC and day.

The manager's perceived in-stock probability is not observed from the data which requires our use of a structural model. Assuming that the decision-maker rationally optimizes a newsvendor problem, the optimal decision vector q^* solves

$$\min_q E\{c_o^T(q-s)^+ + c_u^T(s-q)^+\} \tag{5}$$

where $(x)^+$=max$\{x$,0$\}$ and c_o, c_u are observation-specific overage and underage cost parameter vectors. When the order quantity is larger than the sales required, the local DC experiences overage costs, either through shipping inventory back to the regional DC or through holding inventory in the DC. When the order quantity is smaller than the sales required, the local DC experiences underage costs from increased shipping costs from other DCs. The optimal solution q^* for a specific i, j, t observation satisfies the optimal in-stock probability, also known as the critical fractile,

$$F(q^*) = \frac{c_u}{c_u + c_o} = \frac{1}{1+\gamma} \tag{6}$$

where γ represents the ratio of overage to underage costs, $\gamma = \frac{c_o}{c_u}$, using element-wise division. We do not observe the cost parameter vectors c_o and c_u, and unfortunately these parameters are not separately identified in the data: a DC could ship more inventory either because its underage costs are relatively low, or because its overage costs are relatively high. Note that we can rearrange terms to solve for $\gamma = \frac{1}{F(q^*)} - 1$. If we have an estimate of $F(q^*)$ then we can compute an estimate of γ. From our estimate of demand conditional on the observed promise time, which on average aligns with manager's expected promise time, we can use the functional form of the error on demand to estimate

$$F(q_{ijt}; \beta, \sigma_\epsilon) = P(s_{ijt} \leq q_{ijt}) = \Phi\left(\frac{\log(q_{ijt}) - X_{ijt}\hat{\beta}}{\sigma_\epsilon}\right) \tag{7}$$

Note that our model for demand uses the log of demand, which will predict demand strictly greater than 0. Thus, for observations where the locally sourced quantity is in fact zero, the in-stock probability is exactly 0, which will give us an undefined estimate of the cost ratio. To account for this, we compare those observations with zero fulfillment to the probability that demand is less than .5. In other words, $P(s_{ijt} \leq 0) = \Phi\left(\frac{\log(.5) - X_{ijt}\hat{\beta}}{\sigma_\epsilon}\right)$. This can be interpreted as rounding, where in reality we never observe partial units. In generating the equilibrium we round the optimal quantity and demand to reflect reality, which aligns to our methodology here.

Given our imputed $\hat{\gamma}$ we can estimate the relation to the factors Z_{ijt} that drive the cost ratio. We model the manager's cost ratio with a log-linear relationship to Z_{ijt} as

$$\log(\gamma_{ijt}) = Z_{ijt}\alpha + \xi_{ijt} \tag{8}$$

$$= Z_j\alpha_j + Z_i^{FE}\alpha_i + Z_t^{FE}\alpha_t + \xi_{ijt} \tag{9}$$

where α is a column vector of parameters and ξ_{ijt} is a scalar distributed with mean zero and standard deviation of σ_ξ. We assume ξ_{ijt} is a factor that shifts the cost ratio which is observed by the manager, but unobserved to the researcher. Our cost factors Z_{ijt} include costs related to the SKU such as holding costs or delivery costs from products size or weight, warehouse maintenance costs when the SKU is either ordered frequently in multiple orders ("PercGreater1") or complimentary with other SKUs ("PercMultiOrder"), and fixed effects to the specific DC and day.

As noted earlier, the JD.com data set for the competition does not include the specific inventory held at the end of the day. Because of this, the estimate of the in-stock ratio is biased downwards for observations where the DC completely fulfilled locally for the ijt observation. For these observations, our estimates of γ will be biased upward (which can be seen by examining $\gamma = \frac{1}{F(q^*)} - 1$). One approach would be to simply drop these observations in the analysis, but this would amount to dropping many observations as regional DCs have high service levels. We instead retain observations for the regional DCs to at least give directional insights on the relative cost ratio differences between regional and front DCs. For the sake of our analysis, fortunately front DCs have low service levels and infrequently have 100% fulfillment for an observation, whereas regional DCs frequently have 100% fulfillment. Thus, the DC fixed effect for regional DCs captures the majority of the bias. The costs to local fulfillment for regional DCs is likely lower than our model predicts. Fortunately, our focus is on front DCs not regional DCs, where we expect the fixed effects to be less biased.[5]

Assuming that managers behave with rational expectations by solving a newsvendor problem, we estimate our parameters using the following multi-step procedure:

- Estimate demand parameters $\hat{\beta}$ conditional on the observed promise time in the data
- Impute the optimal in-stock probability $\hat{P}$ from the observed in-stock quantity relative to the manager's expected demand $\hat{\beta}X$

[5]We empirically validated that our results are not heavily influenced by dropping all regional DC observations when estimating the cost side model and the delivery time model. The estimated parameters and the results in the equilibrium remain similar.

– Estimate δ by fitting the observed promise time to the imputed in-stock probability $\hat{P}$ and other factors

– Compute $\hat{\gamma}$ from the the imputed in-stock probability $\hat{P}$; estimate α by fitting $\hat{\gamma}$ to the cost side factors

As mentioned in Olivares et al. (2008), multi-step estimators have additional variance from error in the earlier stages. To account for this, we use bootstrapping to compute the standard errors. We use 100 bootstrap replications, in line with a standard number of 50-200 as pointed out in Efron and Tibshirani (1986). We use maximum likelihood (ML) to estimate the parameters in each stage.[6]

In Appendix: Generating the Equilibrium we discuss the steps taken to simulate our model's predicted equilibrium for comparison to the data, and for simulating equilibrium for counterfactual scenarios.

5 Results

In estimating our model we aggregate the order data to the SKU-DC destination-Day level, as this is the unit of observation defined in the model. Upon aggregating, we have 72,508 unique observations. We first present our results from estimation and then use our model to investigate counterfactual scenarios.

For the sake of brevity we focus on the estimated coefficients that are most relevant to our research question. In Appendix: Interpreting Estimated Coefficients, we more thoroughly explore interpretation of all the estimated coefficients of our model. All three models include day and DC dummy variables to control for average effects from the location and the day. The left section of Table 2 shows that our predictors of demand significantly explain the variation in log sales, with a coefficient of determination (R^2) of .398. In particular, a 1% increase in the median promise time leads to a .104% decrease in sales. As an example, we would expect that on average decreasing the median promise time from two to one day shipping would result in an 5.2% increase in sales. The coefficient is statistically significant at the .01 level, demonstrating that customers are sensitive to the promise time. Therefore, how JD.com makes decisions that impact the delivery of products has important implications for its sales.

[6]For a given model, this is equivalent to estimating the standard deviation of the unobserved component from the uncorrected standard deviation of the residuals in an ordinary least squares (OLS) regression, and using all remaining parameters from the OLS regression.

Table 2: Estimation Results

Demand Model				Promise Time Model				Cost Ratio Model			
X covariates	Estimate			V Covariates	Estimate			Z covariates	Estimate		
Intercept	1.345	(0.0825)	***	*Intercept*	1.450	(0.0153)	***	*Intercept*	1.568	(0.1486)	***
log(MedPromise)	-0.104	(0.0073)	***	*InStockProb*	-0.571	(0.0049)	***	*Attribute1*	0.244	(0.0156)	***
log(AvgPrice)	-0.277	(0.0063)	***	*Type1*	-0.514	(0.0036)	***	*Attribute2*	0.001	(0.0005)	**
Daily Clicks	0.000	(0.0000)	***	*Attribute1*	0.012	(0.0024)	***	*Type1*	-0.482	(0.0176)	***
Type1	0.207	(0.0079)	***	*Attribute2*	0.000	(0.0001)		*PercGreater1*	1.930	(0.0914)	***
PercPlus	-0.055	(0.0393)		σ_ξ	0.355	(0.001)	***	*PercMultiOrder*	-0.145	(0.055)	***
AvgUserLevel	-0.010	(0.013)						σ_ω	2.441	(0.0162)	***
PercMarried	-0.089	(0.0275)	***								
PercFemale	0.101	(0.0201)	***								
AvgEduc	0.036	(0.0136)	***								
AvgCityLevel	0.028	(0.009)	***								
AvgPurchasePower	-0.186	(0.0189)	***								
AvgAge	-0.004	(0.0013)	***								
PercFirstTime	-0.532	(0.2503)	**								
Attribute1	-0.029	(0.0053)	***								
Attribute2	0.004	(0.0002)	***								
σ_ε	0.676	(0.0022)	***								
Observations	72,508			Observations	72,508			Observations	72,508		
R^2	0.398			R^2	0.520			R^2	0.336		

*Notes. The left, middle, right table show the estimates for the demand, cost ratio, and promise time models. All models include day and DC dummy variables. Standard Errors are shown in parentheses, computed with 100 bootstrap replications. ***,**,* denote significance at the 1%, 5%, 10% confidence level.*

Examining the middle section of Table 2 we see that our predictors of promise time significantly explain the variation in log median promise time, with a coefficient of determination of .520. The in-stock probability statistically decreases the median promise time. A higher in-stock probability implies that the firm is less likely to ship from a regional DC, decreasing the delivery time. We would expect that on average increasing the in-stock probability from 0% to 100% decreases the relative promise time by 57.1%. Combining with the previous example, if the customer-perceived promise time was previously two days at a 0% in-stock probability, increasing the in-stock probability to 100% would roughly decrease promise time by 50%, leading to a 5.2% increase in sales. Note that increasing the in-stock probability is substantially harder when the critical ratio is large because the in-stock probability is non-linear in the cost ratio. This implies a difficulty in improving DCs with low service levels: the gain in sales from improving the service level may be less than the relative cost to achieve a higher service level.

In the right panel of Table 2, we see that our predictors of local fulfillment costs significantly explain the variation in the log of our imputed cost ratio, with a coefficient of determination of .336. Recall that a decrease in the cost ratio results in more local fulfillment, either from overage costs decreasing (holding costs or cost to replenish each day) or underage costs increasing (shipping costs increasing from further DCs). For example, we see that the cost ratio decreases for Type 1 SKUs. Likely JD.com has comparative advantages in holding inventory locally, such as better organization or generalized training.

Figure 5 plots the fixed effect for each SKU's impact on the log cost ratio. First we can see that the regional DCs have relatively lower cost ratios. This could be due to more frequent or larger truck deliveries, or lower adminstrative costs for warehouse space, reducing the cost of overage. In particular we notice three DCs with very high coefficients on the cost ratio. These DCs are 16, 17, and 27. From Figure 3 these DC have the lowest average backup fulfillment times, suggesting lower underage costs than other DCs, generating higher cost ratios. DC 16 and DC 17 never fulfill locally, and these DCs are ones that JD.com could consider investing in to decrease the overage cost, thereby decreasing the cost ratio. DC 27 is interesting because it does fulfill locally. Our model predicts that relative to the amount demanded of DC 27, it fulfills relatively less locally than other DCs. Interestingly, this gives DC 27 a larger coefficient for the cost ratio than other DCs that do not have local fulfillment in our data set. In general, for those DCs that we observe in the data that frequently use backup fullfillment, but have relatively lower coefficients on the cost ratio, our model would justify this behavior through lower expected demand.

Figure 5: Estimated Impacts to Critical Ratio by DC

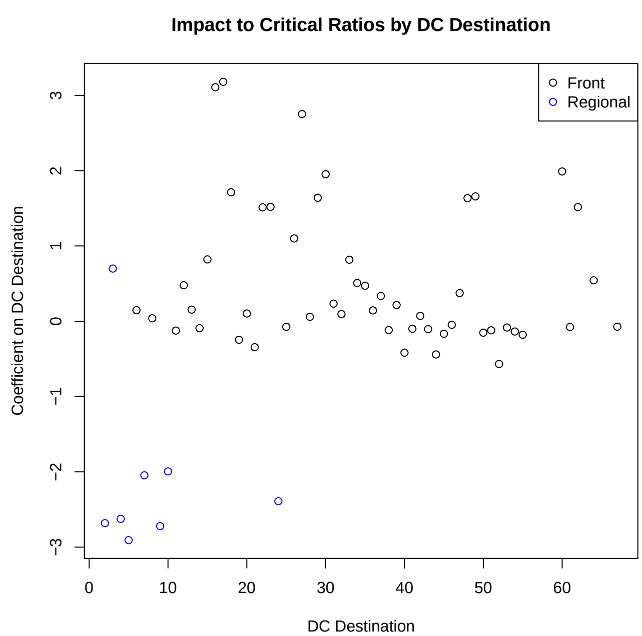


Table 3 gives a comparison of the average equilibrium generated from our model and what we observe in the data. To generate the average equilibrium, we generate 250 simulations by taking random draws of the random variables defined in our model. We use estimates of the parameters estimated in Table 2. From Table 3, we see that our model reasonably maps onto the data observed for JD.com. Our median cost ratio is in-line with the benchmark of 3.0 referenced in Perakis et al. (2020). Instead of assuming a benchmark, we are able to estimate the unobserved cost ratios, an additional insight that our model provides. A cost ratio of 1.0 would imply that overage and underage costs are valued equally, or $c_o = c_u$. Our median cost ratio is substantially above 1.0, so for a typical observation the cost of fulfilling locally is higher than using backup fulfillment. This supports the fact that the percent of quantity fulfilled locally is substantially below 100%.

Table 3: Comparison of Predicted and Observed Equilibrium

	Predicted	Observed
Average Sales Per Observation	3.02	3.48
Total Revenue (in RMB)	20,990,413	22,628,283
Average Revenue Per Observation (in RMB)	289.49	312.08
Median Delivery Time	1.87	2.00
Percent Quantity Filled Local	0.55	0.63
Median Cost Ratio	2.36	

Now we examine the extent to which increasing or decreasing fulfillment from front DCs impacts JD.com. We explore the impact of changing local fulfillment costs by changing the dummy variable parameter on each front DC, α_i. In our model, setting $\alpha_i^{(k)} = \hat{\alpha}_i + \log(k)$ will change the expected cost ratio proportionally by k for a given observation so that $\gamma_{ijt} = k\hat{\gamma}_{ijt}$. This relationship comes from our log-linear model, as

$$Z_j\hat{\alpha}_j + +Z_i^{FE}\alpha_i^{(k)} + Z_t^{FE}\hat{\alpha}_t = Z_j\hat{\alpha}_j + Z_i^{FE}\hat{\alpha}_i + \log(k) + Z_t^{FE}\hat{\alpha}_t \tag{10}$$

$$= \exp(\log(k) + Z_{ijt}\hat{\alpha}) \tag{11}$$

$$= k\hat{\gamma}_{ijt} \tag{12}$$

In other words, we can change the expected cost ratio proportionally through the specific costs to the local DC. Since $\gamma = \frac{c_o}{c_u}$, our approach is equivalent to a proportional change to c_o, e.g. halving overage costs, a proportional change to c_u, e.g. doubling underage costs, or some combination. We can interpret this as proportional effects from improving DC-specific local fulfillment costs through enhancements such as capacity upgrades, staff changes, number of scheduled trucks each day, or technology improvements. Note that a proportional change of $k = 0$, or reducing local fulfillment costs to zero to achieve 100% local fulfillment, implies that $\alpha_i^{(k)} \to -\infty$. A proportional change of $k = \infty$, or increasing local fulfillment costs to be so large that no local fulfillment occurs, implies $\alpha_i^{(k)} \to \infty$.

Table 4 presents the results of our counterfactuals, examining the impacts on fulfillment, delivery time, and revenue. The far left column gives the proportional change, k, from the counterfactual.

Note that $k = 1$ represents the predictions from the original estimated parameters in the model. The left table examines across the data for front DCs only, whereas the right table examines across the data for all DCs. First we discuss the theoretical benefits from front DCs, comparing between $k = 0$, $k = 1$, and $k = \infty$; then we discuss the practical benefits from short-term changes to front DC local fulfillment costs.

Table 4: Counterfactual Scenarios on Local Fulfillment Costs at Front DCs

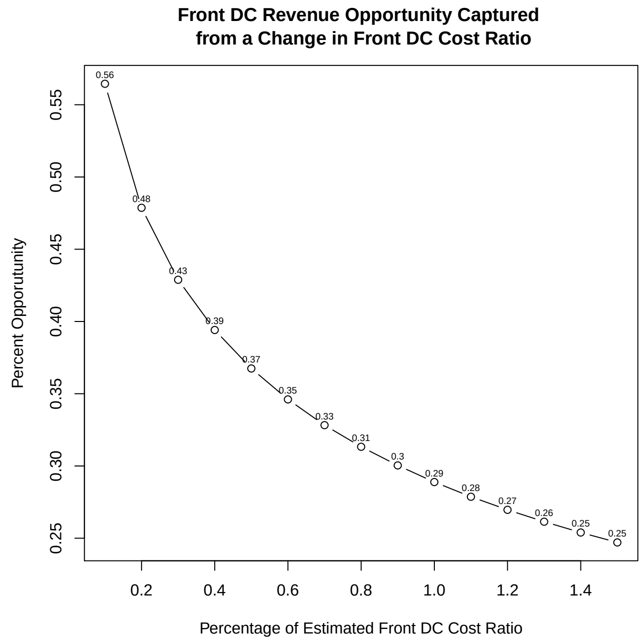
	Front DCs Only			All DCs (including Regional)		
Proportion of $\hat{\gamma}_{FDC}$	Median Delivery Time	Revenue (millions RMB)	Percent Filled Local	Median Delivery Time	Revenue (millions RMB)	Percent Filled Local
$k = 0$	1.35	14.136	1.00	1.40	21.570	0.92
$k = 0.1$	1.70	13.782	0.67	1.66	21.215	0.70
$k = 0.2$	1.79	13.712	0.60	1.72	21.145	0.66
$k = 0.3$	1.85	13.671	0.56	1.76	21.104	0.63
$k = 0.4$	1.89	13.643	0.53	1.78	21.076	0.61
$k = 0.5$	1.92	13.621	0.51	1.81	21.054	0.60
$k = 0.6$	1.94	13.604	0.49	1.82	21.037	0.58
$k = 0.7$	1.96	13.589	0.47	1.84	21.022	0.57
$k = 0.8$	1.98	13.577	0.46	1.85	21.010	0.56
$k = 0.9$	2.00	13.567	0.44	1.86	21.000	0.56
$k = 1$	2.01	13.557	0.43	1.87	20.990	0.55
$k = 1.1$	2.02	13.549	0.42	1.88	20.982	0.54
$k = 1.2$	2.03	13.542	0.41	1.89	20.975	0.54
$k = 1.3$	2.04	13.535	0.40	1.89	20.968	0.53
$k = 1.4$	2.05	13.529	0.40	1.90	20.962	0.53
$k = 1.5$	2.06	13.523	0.39	1.91	20.956	0.52
$k \to \infty$	2.39	13.322	0.00	2.14	20.755	0.27

JD.com recognizes meaningful improvements from its use of front DCs, by examining when $k \to \infty$ changes to $k = 1$, where the median delivery time decreases by 16% resulting in an increase of revenue of 1.8%. On the other hand, when able to completely source demand locally so that $k = 1$ changes to $k = 0$, median delivery decreases by 33% resulting in an increase of revenue of 4.3%. Intuitively, we would expect larger gains from 100% fulfillment because front DC local fulfillment is currently below 50%. Less intuitively, our model captures increasing returns to scale from improved local fulfillment. We discuss this in more detail shortly. Examining the effects across all DCs, the impacts are less pronounced but still meaningful, as revenue from front DCs is a sizable portion of total revenue. In aggregate, when increasing local fulfillment at front DCs there is a positive impact to delivery time of 25% resulting in a positive impact to sales of 2.8%. Clearly, the role of local fullfillment at front DCs is impactful to JD.com's revenue generation, with a large total potential from satisfying all local demand at the front DC instead of through backup fulfillment. Our results show a total potential improvement in regions with local demand for front DCs, so that $k \to \infty$ changes to $k = 0$, of a 44% reduction in median promise time resulting in a 6.1% increase in front DC revenue. JD.com currently only captures 29% of these potential revenue gains from leveraging its front DCs, suggesting a substantial upside for capturing the current gap on missed local fulfillment.

However, our model also demonstrates that there are cost-side reasons for why JD.com is currently unable to completely fulfill in local DCs, despite meaningful improvements to revenue. Achieving improvements to local fulfillment is costly, so JD.com will need to balance the cost of investment to reduce relative costs of fulfilling locally, relative to demand improvements. Figure 8 plots the percent of total revenue opportunity from front DCs that is captured by proportionally changing local fulfillment costs at front DCs. For example, a 50% reduction in local fulfillment costs would result in 37% of the potential revenue from local fulfillment being captured, an increase from 29% currently captured. We can see that revenue improvements from cost reductions are nonlinear: the largest gains in revenue come from substantial decreases in local fulfillment costs, whereas smaller losses are realized from increasing local fulfillment costs. Mathematically, this is generated by the nonlinearity of the in-stock probability, $\frac{1}{1+\gamma}$, with first derivative $-\frac{1}{(1+\gamma)^2}$. This demonstrates that the larger the benchmark γ, the smaller the reduction in in-stock probability from an increase in γ. In practice we can interpret these nonlinear impacts from opportunity cost: when one option is substantially worse than an alternative, a marginal improvement to its cost still makes the alternative option a much better choice; when one choice is substantially better than an alternative, a marginal improvement to its cost makes it an even better choice. Note that 44% of the revenue opportunity at front DCs comes from reducing local fulfillment costs from 10% to 0%. For reference, we estimate regional DCs to have 10-20 times lower cost ratios, so capturing this additional revenue opportunity could be interpreted as converting a front DC to a regional DC.

Because achieving improvements to local fulfillment is costly, JD.com may want to isolate specific front DCs for improvement. For making long-term investment decisions, JD.com may be interested in which front DCs have the most remaining potential revenue gains from enhancements, and which front DCs generate the most negative impacts to revenue from closure. These insights could influence long-term decisions. The left panel of Figure 7 plots the increase in revenue, relative to JD.com's current state, from having 100% fulfillment at each DC, using $k = 0$. We can see that DCs 16, 17, 27 would have the largest gains from improved local fulfillment. Recall from earlier

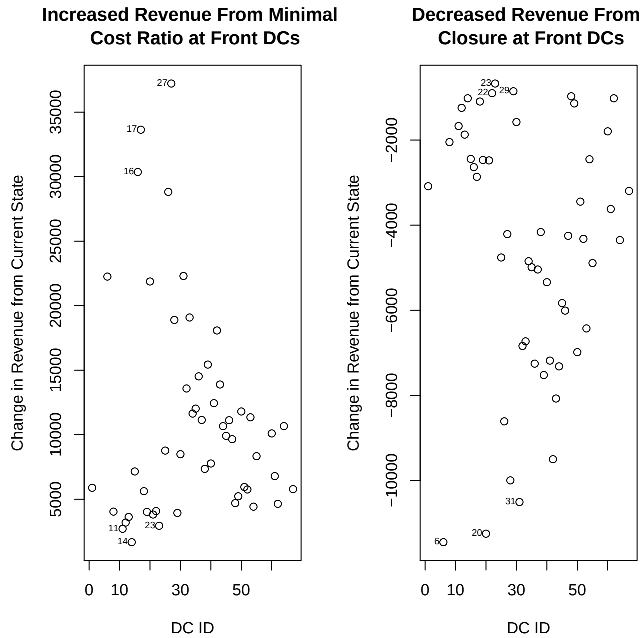
Figure 6: Front DC Revenue Opportunity Captured from Changing Front DC Costs



that DC 16 and 17 both have no local fulfillment in JD.com's data, and DCs 16, 17, and 27 also represented the highest cost ratios in Figure 5. Our model captures the fact that these DCs must have large local fulfillment costs to justify not capturing more demand locally. We can see that DCs 11, 14, and 23 would have the lowest gains from improved fulfillment. These are DCs that already have some local fulfillment and low demand. The right panel of Figure 7 plots the decrease in revenue, relative to JD.com's current state, from having 0% fulfillment at each DC, using $k \to \infty$. We can interpret this as the impact to sales from closing a front DC from fulfilling demand locally. We can see that DC 6, 20, and 31 have the largest negative impacts to sales from closure. In JD.com's data, these are local DCs that currently have high local fulfillment (above 50%) and meaningful sales (above 5000 units). On the other hand, DCs 22, 23, and 29 would have relatively low impacts from closure, as our model predicts these DCs to have low local fulfillment rates and low demand. Note that our model predicts that closure of any front DC would have negative impacts to revenue. Our model predicts that all DCs would carry some inventory, albeit very little for several DCs and none for many SKUs.

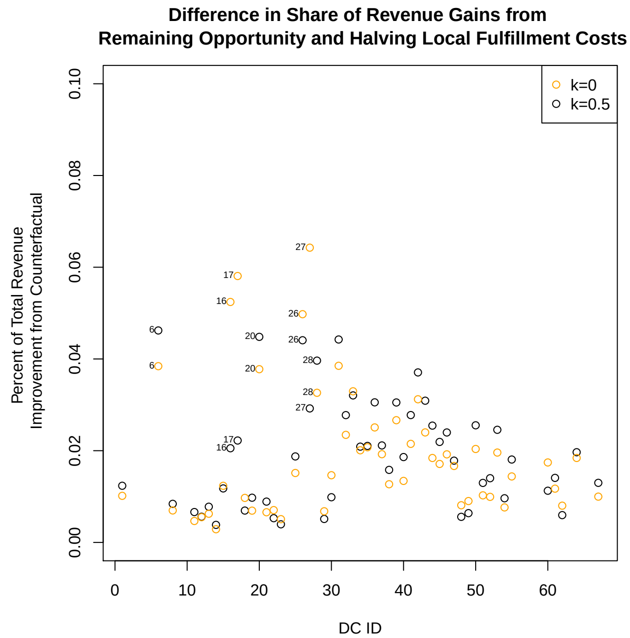
For making short-term investment decisions, JD.com should evaluate the return on investment from a cost reduction relative to revenue improvements. Again, because achieving improvements to local fulfillment is costly, JD.com may want to isolate specific front DCs for short-term investment.

Figure 7: DC Contribution to Revenue from Increasing or Decreasing Cost Ratios



For example, by reducing local fulfillment costs by half, or setting $k = .5$ for specific DCs, which front DCs would experience the largest revenue gains? We examine this question in Figure 8. Figure 8 plots in black the percentage in revenue gained by a specific DC when local fulfillment costs are halved, relative to the total gained across when local fulfillment costs are halved for all front DCs. In orange are the values from the left panel of Figure 7 as a percentage of total revenue gained when local fulfillment costs are zero. Interestingly, the top three local DCs for short-term investment are not the same top three local DCs for long-term investment. A naive approach might invest in DCs 16, 17, 27 from the analysis in the left panel of Figure 7. However, taking into account the practical costs of short-term investment, none of these three would be good candidates to invest now, as demonstrated by how much substantially lower their black points are relative to other black points in Figure 8. Instead, DCs 6, 20, 26 are the best three candidates for short-term investment. Notably, DCs 6 and 20 are in the right panel of 7 for the most costly to close for local fulfillment. Note also that none of these three DCs are the highest in revenue, nor the highest in leveraging current fulfillment. Our model captures the nuanced trade-offs in improving revenue relative to the costs of local fulfillment.

Figure 8: Relative Differences in Returns from DC Investments



6 Conclusion

This paper presents empirical support for JD.com's focus on improved logistics to improve delivery time for its customers. JD.com's customers are sensitive to delivery time, and the front distribution centers improve on delivery time by being closer to the customer. However, we observe in JD.com's data that front distribution centers are underleveraged in their ability to fulfill demand locally, despite having local sales representing a large portion of JD.com's total sales. In turn, the lower service level of front DCs mitigates the improved delivery time of front DCs, reducing the gain in revenue from using front DCs to reach local customers. We build a structural model to capture the costs behind why local DCs are unable to completely fulfill their local demand on their own. We propose that the manager of each distribution center solves a daily newsvendor problem of deciding how many units to fulfill of that day's expected demand, before demand is realized. The manager balances the costs of preparing with too much inventory for the day and shipping inventory from backup DCs when the local units cannot fulfill demand. Our model estimates relative differences in costs across DCs, SKUs, and specific days. These estimated costs could be useful as inputs to inventory allocation algorithms such as Perakis et al. (2020). Using our model, we show that if JD.com invested in front DCs to completely fulfill local demand, median delivery time for those DCs would improve by 33% leading to an increase in their revenue of 4.3%. Across all DCs, this would amount to a 25% improvement in median delivery time and a 2.8% improvement in revenue. We also examine the total potential benefits from using front DCs to get delivery closer to the customer. When fully utilized in areas without a front DC, front DCs have the potential to improve local median delivery times by 44% and improve revenue by 6.1%. We hope these results promote strategic insights into examining new areas to open DCs.

The gains from fully leveraging the front DCs are meaningful, but will require investment such as reducing the cost of daily truck delivery or holding costs. Unfortunately, the largest gains in revenue come from substantial cost reduction; for example, reducing local fulfillment costs from 100% to 50% only improves revenue opportunity captured from 29% to 37%. Still, JD.com may be interested in identifying which front DCs are best for long-term and short-term investment. We identify which front DCs would improve revenue the most from long-term decisions, but show that these are not necessarily the best for short-term decisions. We hope that identifying these opportunities helps JD.com in its ultimate mission to improve demand by providing excellent delivery service to its customers.

To the best of our knowledge, this is the first work to empirically examine the managerial decision of fulfilling demand locally or leveraging backup fulfillment as it impacts customer delivery time in a two-tier distribution network. Future work could improve on some of the limitations of our study. First, the data competition does not provide inventory data, which limits our ability to capture dynamics of ordering more inventory than demand. We assume daily inventory replenishment is primarily through trucks, not through holding inventory. Similarly, our estimation is limited because we assume that the fulfilled quantity observed in the data is the ordered quantity. Second, our work abstracts away from a different challenge of stocking the network as a whole to minimize backup fulfillment. This problem has gained more recent traction in the OM community though papers such as Chen and Graves (2020) and Acimovic and Graves (2015). We have abstracted away from the routing problem for backup fulfillment. Another paper challenge at JD.com (see Yuan and Jing (2018) for details) examines this problem, and this could be a fruitful direction for academic research. Finally, the strategic decision of where to place front DCs also seems promising. One notable structural paper, Holmes (2011), examines where to place Walmart distribution centers, but the fulfillment impacts are different for brick-and-mortar and online retailers. Combining any of these with our current work could improve the richness of our model in improving JD.com operations.

References

Acimovic, J. and Graves, S. C. (2015). Making better fulfillment decisions on the fly in an online retail environment. *Manufacturing & Service Operations Management*, 17(1):34–51.

Akşin, Z., Ata, B., Emadi, S. M., and Su, C.-L. (2013). Structural estimation of callers' delay sensitivity in call centers. *Management Science*, 59(12):2727–2746.

Allon, G., Federgruen, A., and Pierson, M. (2011). How much is a reduction of your customers' wait worth? an empirical study of the fast-food drive-thru industry based on structural estimation methods. *Manufacturing & Service Operations Management*, 13(4):489–507.

Chen, A. I. and Graves, S. C. (2020). Item aggregation and column generation for online-retail inventory placement. *Manufacturing & Service Operations Management*, 0(0):null.

Cui, R., Li, M., and Li, Q. (2019). Value of high-quality logistics: Evidence from a clash between sf express and alibaba. *Management Science*, 66(9):3879–3902.

Efron, B. and Tibshirani, R. (1986). Bootstrap methods for standard errors, confidence intervals, and other measures of statistical accuracy. *Statistical Science*, 1(1):54–77.

Fan, Y. (2018). How e-commerce is transforming rural China. *The New Yorker*.

Fisher, M. L., Gallino, S., and Xu, J. J. (2019). The value of rapid delivery in omnichannel retailing. *Journal of Marketing Research*, 56(5):732–748.

Holmes, T. J. (2011). The diffusion of walmart and economies of density. *Econometrica*, 79(1):253–302.

Ma, C., Lu, J., and Yuan, R. (2018). The secret behind JD.com's super fast delivery. *JD Technology Blog*.

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

Olivares, M., Terwiesch, C., and Cassorla, L. (2008). Structural estimation of the newsvendor model: An application to reserving operating room time. *Management Science*, 54(1):41–55.

Perakis, G., Singhvi, D., and Spanditakis, Y. (2020). Leveraging the newsvendor for inventory distribution at a large fashion e-Retailer with depth and capacity constraints. *Working paper*.

Porteus, E. (2002). *Foundations of Stochastic Inventory Theory*.

Reiss, P. and Wolak, F. (2007). Chapter 64 structural econometric modeling: Rationales and examples from industrial organization. *Handbook of Econometrics*.

Terwiesch, C., Olivares, M., Staats, B. R., and Gaur, V. (2020). OM forum - review of empirical operations management over the last two decades. *Manufacturing & Service Operations Management*, 22(4):656–668.

Yuan, R. and Jing, L. (2018). Highlights from the global optimization challenge. *JD Technology Blog*.

Zhu, F. and Sun, S. (2019). JD: Envisioning the future of retail. *Harvard Business School Case 618-051*.

A Appendix: Generating the Equilibrium

In generating the equilibrium for a set of parameters $\theta = \{\alpha, \beta, \delta\}$, we do the following multi-step procedure for each observation for a given simulation r:

- Generate normal random variables ϵ^r_{ijt}, ω^r_{ijt}, ξ^r_{ijt} with standard deviations $\hat{\sigma}_\epsilon, \hat{\sigma}_\omega, \hat{\sigma}_\xi$ respectively.
- Compute the log cost ratio $\gamma^r_{ijt} = Z^r_{ijt}\alpha + \xi^r_{ijt}$
- Compute the optimal in-stock probability $P^r_{ijt} = \frac{1}{1+\exp(\gamma^r_{ijt})}$
- Compute the optimal order quantity:
  - Compute the log expected promise time percentage:
    $\hat{p}^r_{ijt} = V_{ijt}\delta + (P^r_{ijt} - InStockProb_{ijt})\delta_{InStockProb}$
  - Compute the log expected demand:
    $\hat{s}^r_{ijt} = X_{ijt}\beta + (\hat{p}^r_{ijt} - p_{ijt})\beta_{prom}$
  - Compute the optimal order quantity based on expectations:
    $q^{*,r}_{ijt} = \exp(\hat{s}^r_{ijt} + \hat{\sigma}_\epsilon \Phi^{-1}(\frac{1}{1+\exp(\gamma^r_{ijt})}))$
- Compute actual demand:
  - Compute realized log promise time percentage $p^r_{ijt} = \hat{p}^r_{ijt} + \omega^r_{ijt}$
  - Compute realized demand $s^r_{ijt} = X_{ijt}\beta + (p^r_{ijt} - p_{ijt})\beta_{prom} + \epsilon^r_{ijt}$

B Appendix: Interpreting Estimated Coefficients

In the left panel of Table 2, we see that many of our indicators are significant in predicting log sales. The following increase log sales: daily clicks (significant at .01 level), whether the SKU is carried by JD.com as a Type 1 SKU (significant at .01 level), the percentage of clicks that are from female customers (significant at the .01 level), the average education of interested customers (significant at .01 level), and the average city level of interested customers (significant at the .01

level), and attribute2 on the SKU. We would expect more clicks to lead to more sales because more customers are interested in the product. JD.com prides itself on high-quality products, which is supported by a positive impact of whether the SKU is Type1. Female customers have gained a larger share of JD.com purchases, justifying why more female customers increases sales. The effect of education positive, possibly due to an increased propensity to purchase online products from educated customers. City level is positive, possibly demonstrating the advantages to customers in more rural areas of using JD.com (see Fan (2018)). Attribute2 increases sales, possibly demonstrating that this is an add-on attribute to the main characteristics of the product. The following decrease sales: the median promised time (significant at .01 level), the average price (significant at .01 level), percent of clicks that are plus customers (insignificant at .1 level), the average user level of interested customers (insignificant at the .1 level), percent of clicks from married customers (significant at .01 level), the average purchase power of customers clicking on the SKU that day (significant at .01 level), the average age of of customers clicking on the SKU that day (significant at the .01 level), the percent of clicks that are first time customers (significant at the .05 level), and attribute1 (signficant at the .1 level). As noted in other literature, customers react negatively to increases in promised delivery time and the average price. Married customers may be less likely to purchase if they are likely to go to brick-and-mortar stores regularly for family purchases. Average purchase power decreases sales, demonstrating that conditional on price, customers with more money are more conservative in spending habits. Increases in age decrease sales, as these customers may be more frequent customers at brick-and-mortar stores. More first time customers decreases sales, demonstrating that loyal customers are more likely to purchase. Finally, attribute1 decreases sales. If attribute1 is a primary attribute of the product such as size or weight, customers may be more likely to purchase in stores than online. Note, that all of these effects are conditional on the other variables in the model.

In the middle panel of Table 2, we see that attribute 1 (significant at .01) and attribute 2 (insignificant at .10) decrease promise time. If attribute1 is a primary attribute of the product such as size or weight, then it may increase the cost of holding in inventory, whereas the effect of attribute 2 is less direct. We see that the in-stock probability (significant at .01) and whether the SKU is owned by JD.com reduce the promise time. As mentioned throughout our paper, improved in-stock probability improves the promise time by shipping from closer to the customer. It is possible that Type1 SKUs have a different routing to customers or other differential shipment that improves their promised delivery time.

In the right panel of Table 2, we see that the cost ratio increases from attribute 1 (significant at the .01 level), attribute 2 (significant at the .05 level), and the percentage of orders the SKU has more than one order quantity (significant at the .01 level). It is possible that attribute1 is costly to hold in the warehouse due to size or weight constraints, and similar reasons might cause attribute 2 to increase the cost ratio. We expect that when the percentage of order quantity is greater than one for an SKU, it might be more costly to hold multiple units in the warehouse than the relative loss in total sales from shipping from a regional DC. Lost sales are mitigated by shipping both units, whereas holding costs relative to other inventory are fully realized. We see that the cost ratio decreases from Type 1 (significant at the .01 level), and the percentage of the time that the SKU is on multiple orders with complimentary products (significant at the .01 level). Likely JD.com has comparative advantages in holding inventory locally, such as better organization or generalized training. Unlike for when the quantity is greater than one, it is possible that SKUs that are ordered across complimentary products are less costly to fulfill locally when one of the complimentary products is a product that would be held anyways.