Supplemental Appendix

Nexus Tax Laws and Economies of Density in E-Commerce: A Study of Amazon's Fulfillment Center Network

Jean-François Houde
UW-Madison & NBER

Peter Newberry
The University of Georgia

Katja Seim
Yale University & CEPR & NBER

November 14, 2022

A Appendix: Model

A.1 CES Spending Derivation

Here, we describe the derivation of the spending equations in the CES demand model. The utility of the representative household in county i in year t is given by:

$$U_{it}(q_{i0t},..q_{i3t}) \quad = \quad \left(\sum_{k=0}^{3} \int v_{ikt}(\omega)^{1/\sigma} q_{ikt}(\omega)^{\frac{\sigma-1}{\sigma}} dF_{kt}(\omega) \right)^{\frac{\sigma}{\sigma-1}}$$

where $q_{ikt}(\omega)$ is the quantity of product variety ω purchased from shopping mode k in year t.

The household solves:

$$\begin{aligned} \max_{q_{i0t},..q_{i3t}} \quad & U_{it}(q_{i0t},..q_{i3t}) \\ \text{s.t.} \quad & \sum_{k=0}^{3} \int \tilde{p}_{ikt}(\omega) q_{ikt}(\omega) dF_{kt}(\omega) \leq B_{it} \end{aligned}$$

resulting in the optimal expenditure:

$$e_{ikt} \quad = \quad \int v_{ikt}(\omega) \tilde{p}_{ikt}(\omega)^{1-\sigma} P_{it}^{\sigma-1} B_{it} dF_{kt}(\omega)$$

The term P_{it} represents the Dixit-Stiglitz price index: $P_{it} = \left(\sum_{k=0}^{3} \int v_{ikt}(\omega) \tilde{p}_{ikt}(\omega)^{1-\sigma} dF_{kt}(\omega) \right)^{\frac{1}{1-\sigma}}$.

Given the pricing assumptions and the separability of the marginal utility, expenditures can be rewritten as:

$$\begin{aligned} e_{ikt} \quad &= \quad \int \alpha_{ikt} \omega (\rho_{kt} \omega (1+\tau_{ikt}))^{1-\sigma} P_{it}^{\sigma-1} B_{it} dF_{kt}(\omega) \\ &= \quad \alpha_{ikt} \rho_{kt}^{1-\sigma} (1+\tau_{ikt}))^{1-\sigma} P_{it}^{\sigma-1} B_{it} \int \omega^{2-\sigma} dF_{kt}(\omega) \end{aligned}$$

Taking the log of the relative spending in county i to the offline option, results in Equation 2 from the main text:

$$\begin{aligned} \tilde{e}_{ikt} \quad &= \quad \ln e_{ikt} - \ln e_{i0t} \\ &= \quad \ln(\alpha_{ikt}) - ln(\alpha_{i0t}) + (1-\sigma)(\ln(\rho_{kt}) - \ln(p_{i0t})) \\ & \quad\quad + (1-\sigma) \underbrace{(\ln(1+\tau_{ikt}) - \ln(1+\tau_{i0t}))}_{\Delta\tau_{ikt}} + \underbrace{\ln\left(\int \omega^{2-\sigma} dF_{kt}(\omega) \right)}_{\text{Variety}} \\ &= \quad \xi_{kt} + \lambda_k Z_{it} + \gamma_k C_{it} + \bar{\xi}_i + \Delta\xi_{ct} + (1-\sigma)\Delta\tau_{ikt} + \epsilon_{ikt} \end{aligned}$$

where the final equality comes from the parameterization of tastes for mode k relative to mode 0 and the relative prices:

$$\ln(\alpha_{ikt}) - \ln(\alpha_{i0t}) + (1-\sigma)(\ln(\rho_{kt}) - \ln(p_{i0t})) + \ln\left(\int \omega^{2-\sigma} dF_{kt}(\omega) \right) = \xi_{kt}^{u} + \lambda_k Z_{it} + \gamma_k C_{it} + \bar{\xi}_i + \Delta\xi_{ct} + \epsilon_{ikt}$$

A.2 Calculating Compensating Variation

Here, we describe how we calculate the compensating variation from going from the nexus tax regime to the non-discriminatory regime. Note that we have not estimated all the parameters of the CES model, so we must rely on an approximation of the utility function to calculate the change in consumer welfare. In particular, the distribution of product quality enters the demand estimates only through mode/year fixed-effects, and we do not observed the joint distribution of quantity and prices for each variety. We therefore assume that each mode sells a single variety with a time-varying product quality index α_{ikt}, which leads to the following utility function:

$$U_{it}(q_{i0t},..,q_{i3t}) = \left(\sum_{k=0}^{3} \alpha_{ikt}^{\frac{1}{\sigma}} q_{ikt}^{\frac{\sigma-1}{\sigma}}\right)^{\frac{\sigma}{\sigma-1}}$$

Optimal consumption is then:

$$q_{ikt} = \alpha_{ikt}\tilde{p}_{ikt}^{-\sigma} P_{it}^{\sigma-1} B_{it}$$

and indirect utility is:

$$u(\tilde{p}_{i0t},...,\tilde{p}_{i3t},B_{it}) = \left(\sum_{k=0}^{3} \alpha_{ikt}^{\frac{1}{\sigma}} (\alpha_{ikt}\tilde{p}_{ikt}^{-\sigma} P_{it}^{\sigma-1} B_{it})^{\frac{\sigma-1}{\sigma}}\right)^{\frac{\sigma}{\sigma-1}} = P_{it}^{\sigma-1} B_{it} \left(\sum_{k=0}^{3} \alpha_{ikt}\tilde{p}_{ikt}^{1-\sigma}\right)^{\frac{\sigma}{\sigma-1}}$$

The price index is given by:

$$P_{it} = \left(\sum_{k=0}^{3} \alpha_{ikt}\tilde{p}_{ikt}^{1-\sigma}\right)^{\frac{1}{1-\sigma}}$$

Plugging this into the indirect utility results in:

$$u(\tilde{p}_{i0t},...,\tilde{p}_{i3t},B_{it}) = B_{it}\left(\sum_{k=0}^{3} \alpha_{ikt}\tilde{p}_{ikt}^{1-\sigma}\right)^{\frac{\sigma}{\sigma-1}} \frac{1}{\sum_{k=0}^{3} \alpha_{ikt}\tilde{p}_{ikt}^{1-\sigma}} = B_{it}\left(\sum_{k=0}^{3} \alpha_{ikt}\tilde{p}_{ikt}^{1-\sigma}\right)^{\frac{1}{\sigma-1}}$$

Assume that price changes from $\tilde{p}$ to $\bar{p}$ (and P to $\bar{P}$) and original level of the budget is given by $\tilde{B}_{it}$. The compensating variation for consumer i is the difference between the current budget and B_{it}^*, where B_{it}^* is the budget such that the utility under the two price regimes is equal:

$$\tilde{B}_{it}\left(\sum_{k=0}^{3} \alpha_{ikt}\tilde{p}_{ikt}^{1-\sigma}\right)^{\frac{1}{\sigma-1}} = B_{it}^*\left(\sum_{k=0}^{3} \alpha_{ikt}\bar{p}_{ikt}^{1-\sigma}\right)^{\frac{1}{\sigma-1}}$$

or:

$$B_{it}^* = \bar{B}_{it}\frac{\left(\sum_{k=0}^{3} \alpha_{ikt}\tilde{p}_{ikt}^{1-\sigma}\right)^{\frac{1}{\sigma-1}}}{\left(\sum_{k=0}^{3} \alpha_{ikt}\bar{p}_{ikt}^{1-\sigma}\right)^{\frac{1}{\sigma-1}}}$$

The total compensating variation is the sum fo this difference across all consumers:

$$CV = \sum_i B_{it}^* - \bar{B}_{it}$$

For this calculation, we need to back out α_{ikt} for all modes from the estimates of our model, as we do not estimate this separately from prices. Assuming there is only one variety implies that the estimates of the regression are:

$$\hat{\delta}_{ikt} = \ln(\alpha_{ikt}) - \ln(\alpha_{i0t}) + (1-\sigma)(\ln(p_{kt}) - \ln(p_{i0t})) = \hat{\xi}^u_{kt} + \hat{\lambda}_k Z_{it} + \hat{\gamma}_k C_{it} + \hat{\Delta}\xi_{i0t} + \hat{\epsilon}_{ikt}$$

so:

$$\hat{\alpha}_{ikt} = \exp(\hat{\delta}_{ikt} + \ln(\alpha_{i0t}) - (1-\sigma)(\ln(p_{kt}) - \ln(p_{i0t})))$$

where the variety term drops out because there is only one variety per mode. We normalize $\alpha_{i0t} = 1$ so that:

$$\hat{\alpha}_{ikt} = \exp(\hat{\delta}_{ikt} - (1-\sigma)(\ln(p_{kt}) - \ln(p_{i0t})))$$

For each of the online modes, we construct an average transaction price using data from comScore. See Online Appendix A.6 for details. These prices vary over time, but not across locations. Offline prices, on the other hand, are allowed to vary across space and time. We estimate p_{i0t} using a combination of CPI data from the Bureau of Labor Statistics and prices at walmart.com from comScore. Specifically, we take the ratio of the price index on walmart.com in 2006, $\bar{p}_{WM,t=2006}$, to the 2006 national CPI for Urban Consumers less food and energy, $CPI_{USA,2006}$, and assume that it is representative of the ratio of offline prices to the CPI across all counties and all years. Therefore, with data on the CPI for a given county and year, CPI_{it}, we calculate local offline prices as:

$$p_{i0t} = \frac{p_{WM,t=2006}}{CPI_{USA,2006}} * CPI_{it}$$

To determine CPI_{it} we combine the data on the national CPI for Urban Consumers from 1999 to 2018 with Regional CPI data for Urban Consumers from 2018. Specifically, we assume that the value of the CPI for county i in 2018 is equal to the Regional CPI for Urban Consumers in county i's MSA, when available, and its Census region when MSA data is not available. The regional CPI for Urban Consumers in 2018 comes from the Bureau of Labor Statistics. We then assume that the ratio of the county's CPI to the national CPI is constant across time, meaning we can calculate CPI_{it} across all years using the time-series of national CPI data from 1999-2018:

$$CPI_{it} = \frac{CPI_{i,t=2018}}{CPI_{USA,2018}} * CPI_{USA,t}$$

A.3 Order Flow Matrices

In this section, we provide more details on the order flow matrices. Since the availability of orders is independently distributed across locations, the unconditional Origin-Destination (O-D) probability matrix takes the following form:

$$\Omega^u_{l,i}(N_t) = \left[\prod_{l' | d_{il} > d_{il'}} (1 - \phi_t(K_{l't}))\right] \phi_t(K_{lt})$$

where d_{il} is the distance between county i and location l.

The above matrix can lead to unfulfilled orders (i.e. sum of columns less than one), and so we form the *conditional* fulfillment probability O-D matrix:

$$\Omega_{l,i}(N_t) = \frac{\Omega^u_{l,i}(N_t)}{\sum_{l'=1}^{L} \Omega^u_{l',i}(N_t)}. \tag{1}$$

and use it to predict the volume of orders fulfilled by fulfillment center located in l:

$$q_{jt} = \sum_i Q_{it}(N_t)\Omega_{l,i}\frac{k_j}{K_{lt}} \text{ if } m_j = FC$$

The fraction of orders from county i processed by a sortation center in location l is given by:

$$\Omega^{sc}_{l,i}(N_t) = \mathbb{1}(d_{il} \leq 150)\sum_{l'=1}^{L}\Omega_{l,i}(N_t)\mathbb{1}(d_{l'l} \leq 25)$$

where $d_{l'l}$ is the distance between cluster l and cluster l'. Therefore, the volume of orders processed by sortation center located in l:

$$q_{jt} = \sum_i Q_{it}(N_t)\Omega^{sc}_{l,i} \text{ if } m_j = SC.$$

B Appendix: Estimation

B.1 Return Function

Here we define elements of the return function that are not in the text. Again, the hat indicates a function that was estimated in a previous step:

$X^{j,j'}_{vi}$ is the discounted differences in the number of vertically-integrated transactions:

$$X^{j,j'}_{vi} = \sum_{t=t(j)}^{t(j')}\beta^t\left(\hat{Q}^{vi}_t(N_t|\mathbf{a}^0) - \hat{Q}^{vi}_t(N_t|\mathbf{a}^{j,j'})\right)$$

$X^{j,j'}_{p}$ is the discounted difference in the population density (weighted by facility square-footage):

$$X^{j,j'}_{p} = \sum_{t=t(j)}^{t(j')}\beta^t\left(C^{\text{Pop Dens}}_t(N_t|\mathbf{a}^0) - C^{\text{Pop Dens}}_t(N_t|\mathbf{a}^{j,j'})\right)$$

where $C^{\text{Pop Dens}}$ is the non rent portion of fixed cost $F_t(N_t)$: $C^{\text{Pop Dens}}_t(N_t|\mathbf{a}) = \sum_j k_j \text{Pop Density}_{l_j,t}$

B.2 Moment Conditions and Profit Trade-offs

In this section, we expand the discussion of Table V from the text to include preliminary estimates of the vertical integration parameter, θ_{vi}, and the density cost parameter, κ. We repeat the first panel of Table V in Table I below and add two additional panels.

In the first two rows of the second panel, we report the changes in gross profit (Y) and the number of orders that are vertically integrated (X), averaged across swaps that capture the trade-off between vertical integration and taxes/input prices. This trade-off come from the fact that, in order to increase the number of vertically integrate orders, Amazon must open facilities in more urban areas and, thus, pays higher input prices and/or charges sales tax to more consumers. We again hold the other profit components fixed by conditioning on swaps that exhibit small changes in the other variables.

The left side of the panel focuses on swaps where the average tax rate decreases ($\Delta\text{Tax}^{j,j'} > 0$) and the population weighted number of vertically integrated orders ($\hat{X}_{vi} > 0$) decreases. This set of swaps determines the lowest value of the θ_{vi} such that the total cost savings from vertical integration outweighs the lost revenue from charging additional sales tax, making the observed network optimal. The change in gross profits, averaged across these swaps is $-\$30.92$ million. The average change in number of vertically integrated orders for this subset of swaps is 36.22 million. Similar to the example in Equation 12, we calculate the lower bound of θ_{vi} by taking the ratio of these numbers, which equals \$0.85 of cost saving per order. The lower bound is \$0.17 per order if instead we use the input price as our profit shifter. To determine the upper bound, we compute the average changes in gross profit and vertically integrated orders for swaps that feature the opposite side of the trade-off. Taking the ratio of these, we calculate the upper bound as 2.08 or \$0.27 per order, depending on the profit shifter.

In the bottom panel of the table, we focus on the trade-off that identifies the additional rental cost of locating in a densely populated area, κ. The trade-off is that Amazon pays the additional cost of higher density as it locates closer to consumers, reducing the shipping distance. Note that the fact that the shipping distance impacts the shipping costs implies that this trade-off does not identify κ separately from θ_d but, instead, identifies the ratio of these two parameters. In the left side of the panel, we focus on the subset of swaps where the population density decreases ($X_{pop} > 0$) and the total population weighted shipping distance increases ($\hat{X}_d < 0$). Note that there is no hat on X_{pop} because this is a fixed-cost and, thus, is not a function of the number of orders (i.e., it is exogenous). Additionally, we hold the gross profit fixed by conditioning on swaps that have small changes in taxes and input prices. Given a value of θ_d, this set of swaps identifies the highest value of κ such that the cost of having a facility in a more densely populated area does not outweigh the savings from shorter shipping distances, making the observed network optimal.

The average change in population density for this subset of swaps is 24.17 people per square mile (100s) where the average change in shipping distance is -71.97 hundred million miles. Again, similar to the example in 12, we can take the ratio of these to calculate the lower bound of $\frac{\theta_d}{\kappa}$, which is 0.336. The right side of the panel, which focuses on the opposite side of the trade-off, shows that the upper bound of the ratio is 0.307. Notice again that this exercise does not restrict the upper bound to be higher than the lower bound. Using the mid-point of the estimates of θ^d from the top panel of the table (0.17), we determine that κ is approximately \$0.50 per unit of population density.

Table I: Moment Conditions and Profit Trade-offs

(a) Distance trade-offs

	Lower bound: θ_d $Z^{j,j'} = 1(\Delta\text{Shifter}^{j,j'} > 0 \ \& \ \hat{X}_d^{j,j'} < 0)$		Upper bound: θ_d $Z^{j,j'} = 1(\Delta\text{Shifter}^{j,j'} < 0 \ \& \ \hat{X}_d^{j,j'} > 0)$	
	$E(Y\|Z)$	$E(X_d\|Z)$	$E(Y\|Z)$	$E(X_d\|Z)$
ΔShifter (Gross Profit)				
(a) Tax	-13.30	-93.60	37.10	140.19
(b) Input prices	-5.94	-82.68	7.60	128.21
Bounds: $\frac{E(Y\|Z)}{E(X_d\|Z)}$				
(a) Tax	0.14		0.26	
(b) Input prices	0.07		0.06	

(b) Vertical integration trade-offs

	Lower bound: θ_{vi} $Z^{j,j'} = 1(\Delta\text{Shifter}^{j,j'} > 0 \ \& \ \hat{X}_{vi}^{j,j'} > 0)$		Upper bound: θ_{vi} $Z^{j,j'} = 1(\Delta\text{Shifter}^{j,j'} < 0 \ \& \ \hat{X}_{vi}^{j,j'} < 0)$	
	$E(Y\|Z)$	$E(X_{vi}\|Z)$	$E(Y\|Z)$	$E(X_{vi}\|Z)$
ΔShifter (Gross Profit)				
(a) Tax	-30.92	36.22	34.10	-16.36
(b) Input prices	-3.51	20.44	4.58	-16.81
Bounds: $\frac{E(Y\|Z)}{-E(X_{vi}\|Z)}$				
(a) Tax	0.85		2.08	
(b) Input prices	0.17		0.27	

(c) Density/distance trade-offs

	Lower bound: $\frac{\theta_d}{\kappa}$ $Z^{j,j'} = 1(X_{pop}^{j,j'} > 0 \ \& \ \hat{X}_d^{j,j'} < 0)$		Upper bound: $\frac{\theta_d}{\kappa}$ $Z^{j,j'} = 1(X_{pop}^{j,j'} < 0 \ \& \ \hat{X}_d^{j,j'} > 0)$	
	$E(X_{pop}\|Z)$	$E(X_d\|Z)$	$E(X_{pop}\|Z)$	$E(X_d\|Z)$
ΔShifter (Fixed Cost)				
Pop Density	24.17	-71.97	-33.44	108.97
Bounds: $\frac{E(X_{pop}\|Z)}{-E(X_d\|Z)}$	0.336		0.307	

Notes: In selecting swaps for inclusion in each instrument category, we condition on population-weighted tax, input price, and distance changes. The statistics in the body of the table, however, represent order-weighted aggregates. The variable ΔShifter refers to the change in one of two population-weighted profit shifters: taxes and average input prices.

B.3 Network Optimization Algorithm

We describe the simulation-based algorithm we use to solve Amazon's network optimization problem, which is called the Population-Based Incremental Learning (PBIL) algorithm developed by Baluja (1994). The central idea of the algorithm is to convexify the optimization problem by iterating over the *probability* of opening a facility in a given location, rather than on the vector of binary choices. The algorithm starts with a uniform probability of entering each location, and progressively refine the guess by bringing the probabilities closer to one or zero to maximize the expected profit. To avoid converging too quickly to a local optimum, the algorithm perturbs a random number of location choice probabilities. This is the genetic component of the procedure. The algorithm stops when the maximum profit configuration stops evolving. The procedure is then repeated using a different sequence of random numbers (*runs*). As this number goes to infinity, the algorithm converges to a global maximum.

Let N denotes the number of unique locations. The choices is summarized by two $N \times 1$ vectors: $A = \{a_{i,fc}, a_{i,sc}\}_{i=1,\dots,N}$ where $a_{i,fc} \in \mathbb{N}^{+}$ and $a_{i,sc} \in \{0,1\}$. Each location can include more than one fulfillment center, and at most one sortation center. In addition, we impose the restriction that a sortation center cannot be placed in a location without a fulfillment center. The algorithm treats those decisions as stochastic, and iterate over the probability of each network configuration. We use a Poisson probability distribution to describe the choice of fulfillment centers, and a binomial distribution to describe the choice of sortation centers. Let $\Lambda = \{\lambda_{i,fc}, \lambda_{i,sc}\}_{i=1,\dots,N}$ denote a matrix of parameters describing the location choice probabilities: $a_{i,fc} \sim \text{Pois}(\lambda_{i,fc})$ and $\Pr(a_{i,sc} = 1) = \lambda_{i,sc}$. Importantly, we are not restricting the number of facilities. The optimization problem searches for the location and number of facilities of each type.

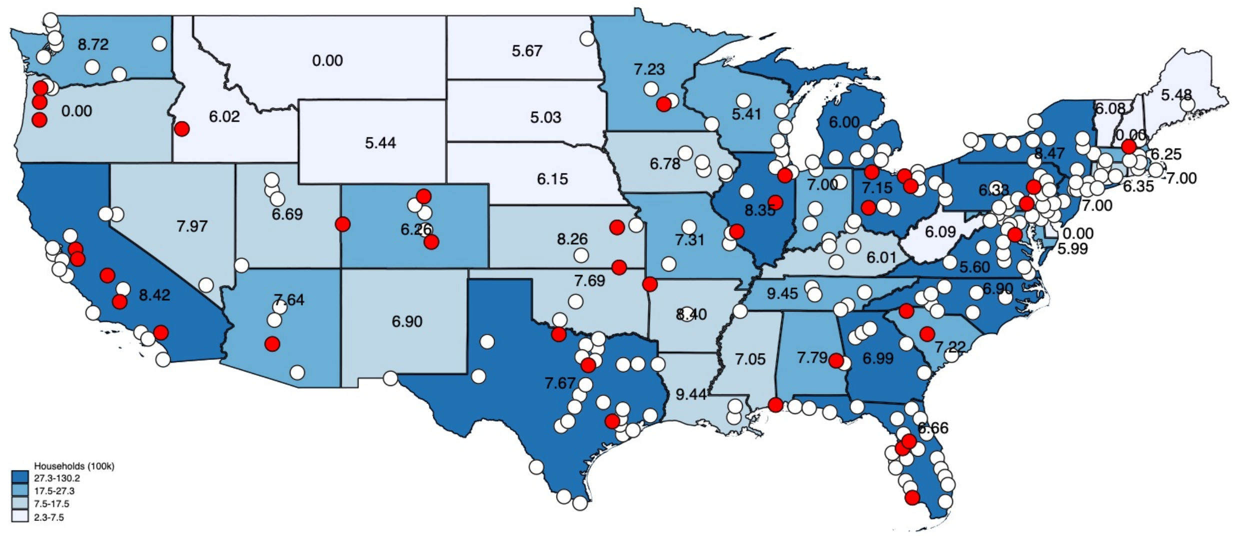
The algorithm proceeds as follows. The starting parameter value Λ^0 is defined such that each location is chosen with uniform probability. We also scale the parameters so that the expected number of locations is equal to the number of observed facilities of each type in a given year. At generation k, the algorithm updates the choice probability vector as follows:

1. *Maximization step:*
   (a) Sample S network configurations from probability distribution Λ^k
   (b) For each s, calculate the flow of orders and the aggregate profits: Π_t^s
   (c) Identify the profit maximizing configuration: $A_k^{\max} = \arg\max_{s\in\{1,\dots,S} \Pi^s$.
2. *Hill-climbing step:* Update network probability parameters using the convex combination of Λ^k and $A_k^{\max}$
$$\Lambda^{k+1} = (1-\alpha)\Lambda^k + \alpha A_k^{\max}$$
3. *Mutation step:*
   (a) For each location/facility type draw $u_{i,j}^k \sim U[0,1]$ and $e_{ij}^k \sim U[0,1]$
   (b) If $u_{i,j}^k < \eta^1$, perturb (i,j) choice probability according to:
$$\lambda_{i,j}^{k+1} = \lambda_{i,j}^k \eta^2 + e_{ij}^k (1-\eta^2)$$
4. If $\Pi^{\max} < \Pi_k^{\max}$ update the profit maximization network:
$$A^{\max} = A_k^{\max} \text{ and } \Pi^{\max} = \Pi^{\max}$$

Steps (1)-(4) are repeated until the network $A^{\max}$ stops changing. In practice we stop the algorithm when $\Pi^{\max}$ has remained constant for $\bar{S}$ generations. To ensure that the algorithm identifies a global maximum, we repeat the stochastic process over a large number of runs, and identify the most profitable network across L runs. Each run yields a potentially different network configuration because the sequence of random numbers is different across generations. When the number of ideal facilities is small (e.g. $t = 1999$ $n^{fc} = 4$) most runs lead to the same configuration, and the stochastic algorithm identifies a unique global maximum. For larger networks, the largest profit network changes from run to run. In these cases, it is best to think of the result as a stochastic approximation of the optimal network.

In Figure 1 we display the locations that are in the choice set (white), as well as the locations that our model predicts in 2018 (red). A total 39 of the 48 contiguous states have a location, with most of those states without a location being on the bottom of the population distribution.

Figure 1: Possible Locations (white) and 2018 Chosen Locations (red)



Notes: States are shaded by quintiles of the number of households with the darker color indicating more households. The average sales tax rate is displayed.

C Appendix: Robustness

C.1 Demand Specification

In this section, we explore the implications of our demand model. One worry is that we are not accounting for unobserved time-varying heterogeneity based on a consumer's location. In the current model, we allow for unobserved growth in preferences for online shopping at the census division level via a set of division-year fixed effects. However, it could be the case that there is geographic variation in growth at the mode level and/or at a different level of spatial granularity. In Table I we present the estimated constant and σ for 5 different models, where each model assumes a different level of fixed effects (indicated at the bottom of the table). Specification (3) is our baseline.

Overall the table demonstrates that the estimate of σ is always negative and significant at the 5% level, but the magnitude is sensitive to the level of fixed effects. Controlling for a finer level of geographic variation appears to be important, as demonstrated by both the estimated constant and σ, and allowing for mode-level unobserved growth results in an estimate below 1 (in absolute value). The reason for this is that the mode-year-level fixed effects absorb much of the across mode and time variation in tax rates used to identify σ. We note that the estimate of the baseline model (3) is the closest to the estimates in the previous literature (e.g., (Einav et al., 2014) and (Baugh et al., 2018)).

Table I: Demand Estimates (Alternative Specifications)

	(1)	(2)	(3)	(4)	(5)
Variable name					
Elas of Subs	-1.187**	-1.189**	-1.516**	-0.830*	-0.958*
	(0.391)	(0.395)	(0.399)	(0.410)	(0.433)
Constant	-3.585	-2.942	-6.378*	-3.642	-7.070*
	(2.768)	(2.857)	(2.987)	(2.858)	(2.989)
Obs	52,488	52,488	52,488	52,488	52,488
R-Sq	0.600	0.604	0.608	0.604	0.609
FEs	None	Region-Year	Division-Year	Region-Year-Mode	Division-Year-Mode

Notes: *** p<0.01, ** p<0.05, * p<0.1. Standard errors in parentheses. The results are from models that are equivalent to the left panel of Table III, but with a different dependant variable. We omit the other parameter estimates for space.

In Table II, we examine how the different demand specifications impact the cost estimates. The specification indicated at the top of the table corresponds to the specifications in Table I. The point estimates are generally similar across the specifications, but the confidence intervals are slightly larger under the baseline model. The estimates of the distance cost in specifications (4) and (5) are about 20% smaller than the baseline, which is in line with the intuition that the tax-distance trade off varies with the magnitude of σ. Similarly, the VI parameter is also slightly smaller in specifications (4) and (5).

Table II: Cost Estimates (Alternative Demand Specifications)

	(2)			(3)			(4)			(5)		
	Est.	CI		Est.	CI		Est.	CI		Est.	CI	
θ_d	0.30	0.24	0.42	0.34	0.26	0.49	0.26	0.22	0.36	0.28	0.22	0.38
θ_{vi}	-0.48	-0.71	-0.09	-0.52	-0.91	0.01	-0.46	-0.63	-0.16	-0.44	-0.70	-0.07
κ	0.85	0.63	1.28	0.98	0.69	1.56	0.74	0.56	1.14	0.78	0.56	1.24
Mom		14			14			14			14	
Ineq		5577			5577			5577			5577	

Notes: Each specification corresponds to a demand model from Table I.

C.2 Alternative Measures of Household Spending

A concern may be that our procedure for constructing online spending (described in A.1) drives our results. We therefore investigate the robustnness of the CES demand estimates from Section 4.1 to the use of alternative dependent variables. Table III reports the results, where the unit of observation across all specification is at the level of the county, year, and shopping mode. The first column uses the log difference between the average spending on a given mode calculated using only the comScore data (i.e., the raw data) and offline retail spending for a given county. In the second column, we replace the raw averages with spending adjusted for extensive margin underreporting through the Forrester online purchase probabilities. The elasticity of substitution is similar for these two specifications, but slightly lower than our main estimate (column 4). The results in column 3, where we use a weighted average of the spending variable using population weights, indicate that this is most likely due to the sampling of households in the comScore data. Last, scaling the data to match reported aggregates in column 4 results in changes to the estimated mode/year FEs (not shown), as the levels of spending change. We also estimate a slightly smaller elasticity of substitution as compared to column (3).

The similar elasticity estimates in (3) and (4) suggest, as was discussed in Section A.1, that the scaling of the data primarily acts to adjust the levels of spending (i.e., the fixed effects), with the cross-sectional variation in market shares largely preserved. We note that if these regressions where done with the levels of spending as the dependent variable rather than log differences in spending, then inflating the data by a constant would not impact the tax coefficient. Therefore, that the tax coefficients differ slightly between columns (3) and (4) is due to the non-linearity of the model. Overall, these robustness exercises give us confidence that the specifics behind the construction of the spending variable are not a significant driver of our results.

Table III: CES Model with Alternative Spending Variables

Variable name	(1)	(2)	(3)	(4)
Elasticity of substitution	-1.295**	-1.331**	-1.703**	-1.516**
	(0.284)	(0.284)	(0.337)	(0.399)
Constant	-4.452*	-5.392*	-3.788	-6.378*
	(2.124)	(2.128)	(2.522)	(2.987)
Obs	52,488	52,488	52,488	52,488
R-Sq	0.513	0.519	0.454	0.608
Spending Variable	Raw	Raw + Forr Adj	Raw + Forr Adj + Pop weights	Raw + Forr Adj + Pop weights + Infl

Notes: *** p<0.01, ** p<0.05, * p<0.1. Standard errors in parentheses. Dependent variable is defined as log of expenditure on each online mode divided by offline expenditure. Specification (1) uses the county average of the raw comScore spending; specification (2) the county average of comScore spending adjusted by the probability of online purchase by demographic group from Forrester; specification (3) the weighted county average of the same spending variable using Census-based sampling weights; and specification (4) further inflates county-level spending to match outside spending statistics. See text for detail.

C.3 Order Flow Model

In this section we compare the estimates from two alternative order flow models estimates. The first two columns in Table IV reproduce our main specification results (MD). Recall that the parameters are estimated by minimizing the distance between the predicted and observed number of employees in 2017 for a sample of sortation center and fulfillment center facilities. The third and fourth columns estimate the model instead by matching aggregate moments capturing the correlation between facility characteristics (population density and capacity) and number of employees. This approach does not exploit the full richness in the distribution of facility sizes, but is more immune to measurement error in reported employees. The last column restricts the stock-out probability to be zero ($\psi \to \infty$). This implies that orders are fulfilled by the nearest cluster.

Table IV: Order Flow Model Parameter Estimates (Alternative Specifications)

(a) Parameter estimates

Parameters	Stockout (MD) Est.	Stockout (MD) SE	Stockout (GMM) Est.	Stockout (GMM) SE	Nearest Est.	Nearest SE
Availability – ψ	0.49	0.21	1.54	1.93	–	
Output – γ	0.47	0.10	0.74	0.10	0.31	0.08
A_{fc}	1.81	0.75	0.68	0.28	3.52	1.00
A_{sc}	0.41	0.17	0.10	0.05	0.68	0.25
Objective func.	37.62		2.24		40.08	

(b) Goodness of fit

Moments	Obs. moments Est.	Obs. moments SE	Model predictions MD	Model predictions GMM	Model predictions Nearest
Regression: Empl (log)					
Intercept	5.39	0.55	6.40	5.39	6.61
FC x Pop dens.	0.25	0.08	0.10	0.25	0.06
FC x Capacity	0.64	0.20	0.57	0.64	0.31
SC	0.21	0.57	-0.81	0.21	-1.01
Annual growth rate	0.24	0.00	0.21	0.24	0.20
2017 Empl. (log)	4.83	0.01	4.92	4.83	4.93

Notes: The model predictions are the point estimate of the goodness of fit regressions, where MD and Nearest are already presented in the main text.

The scale parameter γ is estimated to be less than one in all three specifications (i.e. increasing return to scale). The MD specification (baseline) is in the middle of that range (i.e. 0.47). The main difference across the three specifications is in the availability parameter ψ. Recall that this parameter determines the relationship between the size of a cluster and the probability that an order is fulfilled. The GMM specification identifies this parameter by matching the observed positive relationship between population density around a cluster, and the number of employees in a facility. The estimate suggest a positive stock-out probability ($\psi << \infty$), but less frequent than in our baseline specification (i.e. $0.49 < 1.54$). This estimate is very noisy however, presumably due to the weakness of the reduced-form moments.

Table V: Cost Estimates (Alternative Order Flow Specifications)

	Specification A: MD Order-flow model								
	Est.	CI—		Est.	CI—		Est.	CI	
SC: Distance (x100 m.)	0.16	0.15	0.17	0.59	0.41	0.88	0.34	0.26	0.49
SC: VI Orders							-0.52	-0.91	0.01
FC: Pop. density (x100)				2.06	1.48	3.17	0.98	0.69	1.56
Nb of moments	4.00			8.00			14.00		
	Specification B: GMM Order-flow model								
	Est.	CI—		Est.	CI—		Est.	CI	
SC: Distance (x100 m.)	0.13	0.13	0.00	0.23	0.13	0.37	0.16	0.10	0.22
SC: VI Orders							-0.28	-1.09	-0.15
FC: Pop. density (x100)				0.67	0.33	1.34	0.52	0.30	0.84
Nb of moments	4.00			8.00			14.00		

Notes: The MD order flow model is the same as presented inthe main text. The GGM order flow model corresponds to the MD estimates in Table IV.

Table V compares the estimated cost function under the alternative order-flow estimates. Note that we cannot use the "nearest" specification to construct our moment inequality estimator. This is because the number and size of facilities within a cluster do not impact the variable profit. Without stock-outs, the firm would select a single facility per cluster, which is not consistent with the observed roll-out.

The estimates obtained using the GMM specification are qualitatively similar to our baseline estimates (Panel A), but lead to lower average fulfillment costs.