EC.1. An Alternative Linear Program and Its Fulfillment Heuristic

Here, we will provide an alternative linear program formulation that explicitly takes into account the fulfillment consolidation decision. Let $\sigma_j^q : \mathcal{S}_I \mapsto \mathcal{S}_K$ be as defined in Section 4. Let H_t^π denote the history of realized events and decisions (i.e., the accumulated information) up to the beginning of period t under policy π and let $G_j^{\pi qt}(\sigma_j^q, h) = P(X_j^{\pi qt} = \sigma_j^q \,|\, D_j^{qt} = 1,\, H_t^\pi = h)$ denote the probability of fulfilling order type q from region j during period t with assignment vector σ_j^q under policy π given history $H_t^\pi = h$. (The expression $X_j^{\pi qt} = \sigma_j^q$ is a shorthand for $X_{kij}^{\pi qt} = \mathbf{1}\{\sigma_j^q(i) = k\}$ for all $i \in q$.) By abuse of notation, we will also write $G_j^{\pi qt}(\sigma_j^q) := \mathbf{E}_{H_t^\pi}[G_j^{\pi qt}(\sigma_j^q, H_t^\pi)] = \mathbf{E}_{H_t^\pi}[P(X_j^{\pi qt} = \sigma_j^q \,|\, D_j^{qt} = 1,\, H_t^\pi)] = P(X_j^{\pi qt} = \sigma_j^q \,|\, D_j^{qt} = 1)$, where the expectation is taken over the set of all possible histories under policy π up to the beginning of period t. Observe that we can write:

$$\begin{aligned}
\mathbf{E}\left[X_{kij}^{\pi qt} \,\middle|\, D_j^{qt} = 1\right] &= \sum_{\{\sigma_j^q :\, \sigma_j^q(i) = k\}} G_j^{\pi qt}(\sigma_j^q) \quad \text{and} \\
\mathbf{E}\left[\max_{i\in q}\{X_{kij}^{\pi qt}\} \,\middle|\, D_j^{qt} = 1\right] &= \sum_{\{\sigma_j^q :\, \exists i \in q,\, \sigma_j^q(i) = k\}} G_j^{\pi qt}(\sigma_j^q).
\end{aligned}$$

So, for any policy $\pi \in \Pi$, the objective function in J^* can be written as:

$$\sum_{t,j,q,k} \lambda_j^q \left[\sum_{i\in q} \sum_{\{\sigma_j^q :\, \sigma_j^q(i) = k\}} c_{kij}\, G_j^{\pi qt}(\sigma_j^q) \;+ \sum_{\{\sigma_j^q :\, \exists i\in q,\, \sigma_j^q(i) = k\}} b_{kj}\, G_j^{\pi qt}(\sigma_j^q) \right].$$

Taking an expectation over the constraints in J^* immediately yields the following, where we use J_{ALP} to denote Alternative Linear Program:

$$\begin{aligned}
J^* \;\geq\; J_{ALP} \;:=\; \min_u \quad & \sum_{t,j,q,k} \lambda_j^q \left[\sum_{i\in q} \sum_{\{\sigma :\, \sigma(i) = k\}} c_{kij}\, u_{\sigma j}^{qt} \;+ \sum_{\{\sigma :\, \exists i\in q,\, \sigma(i) = k\}} b_{kj}\, u_{\sigma j}^{qt} \right] \\
\text{s.t.} \quad & \sum_t \sum_j \sum_{q \ni i} \lambda_j^q \left[\sum_{\{\sigma :\, \sigma(i) = k\}} u_{\sigma j}^{qt} \right] \leq S_{ki} \quad \forall k, i && \text{(EC.1)} \\
& \sum_k \sum_{\{\sigma :\, \sigma(i) = k\}} u_{\sigma j}^{qt} = 1 \quad \forall q, t, j, i \in q && \text{(EC.2)} \\
& 0 \leq u_{\sigma j}^{qt} \leq 1 \quad \forall q, t, j, \sigma && \text{(EC.3)}
\end{aligned}$$

For brevity we simply write σ instead of σ_j^q. Per our discussions above, any feasible policy $\pi \in \Pi$ essentially corresponds to a set of distributions $\{G_j^{\pi qt}\}$ over the set of fulfillment assignment $\{\sigma_j^q\}$.

(We implicitly allow π to be a randomized policy.) Since constraints (1)-(3) in J^* imply (EC.1)-(EC.3), this proves that J_{ALP} is a lower bound for J^*. (Indeed, it is not difficult to see that $J^* \geq J_{ALP} \geq J_{DLP}$.)

Constraints (EC.2) and (EC.3) can be further simplified into $\sum_\sigma u^{qt}_{\sigma j} = 1$ and $u^{qt}_{\sigma j} \geq 0$. This is so because, for each triplet (q,i,j) where $i \in q$, the set of fulfillment vectors $\{\sigma^q_j\}$ can be decomposed into $\cup_k \{\sigma^q_j : \sigma^q_j(i) = k\}$. If we now define $c^q_{\sigma j} := \sum_{i\in q} c_{\sigma(i)ij} + \sum_{\{k:\, \exists i \in q,\, \sigma(i)=k\}} b_{kj}$ (it can be interpreted as the cost of applying assignment vector σ to order type q from region j), we can rewrite J_{ALP} in a more compact form as follows:

$$\begin{aligned} J_{ALP} \; := \; \min_u \; & \sum_t \sum_j \sum_q \lambda^q_j \left[\sum_\sigma c^q_{\sigma j}\, u^{qt}_{\sigma j} \right] \\ \text{s.t.} \; & \sum_t \sum_j \sum_{q \ni i} \lambda^q_j \left[\sum_{\{\sigma:\, \sigma(i)=k\}} u^{qt}_{\sigma j} \right] \leq S_{ki} \quad \forall k, i & \text{(EC.4)} \\ & \sum_\sigma u^{qt}_{\sigma j} = 1 \quad \forall q, t, j & \text{(EC.5)} \\ & u^{qt}_{\sigma j} \geq 0 \quad \forall q, t, j, \sigma & \text{(EC.6)} \end{aligned}$$

The linear program J_{ALP} has a natural interpretation. If demands are deterministic and are arriving with rates $\{\lambda^q_j\}$, then the variable $u^{qt}_{\sigma j}$ can be interpreted as the fraction of demand type q from region j during period t fulfilled with assignment vector σ. Let $U^q_{\sigma j}$ denote the number of times order type q from region j is fulfilled using assignment vector σ during the selling horizon. We can formulate the time-aggregate version of J_{ALP} as follows:

$$\begin{aligned} \tilde{J}_{ALP} \; := \; \min_U \; & \sum_j \sum_q \sum_\sigma c^q_{\sigma j}\, U^q_{\sigma j} \\ \text{s.t.} \; & \sum_j \sum_{q \ni i} \sum_{\{\sigma:\, \sigma(i)=k\}} U^q_{\sigma j} \leq S_{ki} \quad \forall k, i & \text{(EC.7)} \\ & \sum_\sigma U^q_{\sigma j} = T\lambda^q_j \quad \forall q, j & \text{(EC.8)} \\ & U^q_{\sigma j} \geq 0 \quad \forall j, q, \sigma & \text{(EC.9)} \end{aligned}$$

Let $\{\mathbf{u}^{qt}_{\sigma j}\}$ and $\{\mathbf{U}^q_{\sigma j}\}$ denote an optimal solution of J_{ALP} and $\tilde{J}_{ALP}$, respectively. (These solutions need *not* be unique.) It is not difficult to see that $J_{ALP} = \tilde{J}_{ALP}$. First, since $U^q_{\sigma j} = \sum_t \lambda^q_j \mathbf{u}^{qt}_{\sigma j}$ is a feasible solution to $\tilde{J}_{ALP}$, we have $\tilde{J}_{ALP} \leq J_{ALP}$. To show the converse, simply note that $u^{qt}_{\sigma j} = \mathbf{U}^q_{\sigma j}/(T\lambda^q_j)$ is a feasible solution to J_{ALP}. Indeed, it is also optimal. This proves our claim. (In the scaled problem, we have $\tilde{J}_{ALP}(\theta) = \theta \tilde{J}_{ALP}$ and $\mathbf{U}^q_{\sigma j}(\theta) = \theta \mathbf{U}^q_{\sigma j}$. So, this claim still holds.)

We define our fulfillment heuristic below.

Probabilistic Consolidation Control (PCC)

```
Input: $u^{qt}_{\sigma j} = \mathbf{U}^q_{\sigma j}/(T\lambda^q_j)$, where $\{\mathbf{U}^q_{\sigma j}\}$ is an optimal solution of $\tilde{J}_{ALP}$
During period $t$, for an order type $q$ from region $j$, do:
    1. Sample $\sigma$ with probability $u^{qt}_{\sigma j}$
    2. For each $i \in q$, do:
        - If $S_{\sigma(i)i} \geq 1$, fulfill item $i$ from facility $\sigma(i)$
        - Otherwise, fulfill item $i$ from facility 1.
```

It is not difficult to show using similar arguments as in the proof of Theorem 1 that $\mathbf{E}[C_{PCC}(\theta)] - J^*(\theta) = O(\sqrt{\theta})$; so, PCC is, in fact, asymptotically optimal. However, as mentioned in Section 1.1, solving $\tilde{J}_{ALP}$ for practical implementation may present a technical challenge. (For each pair (q, j), $\tilde{J}_{ALP}$ has $|\mathcal{S}_K|^{|q|}$ variables; in contrast, $\tilde{J}_{DLP}$ only has $|\mathcal{S}_K|.|q| + |\mathcal{S}_K|$ variables.) Perhaps there is a way to solve $\tilde{J}_{ALP}$ efficiently; we leave this as future research project.

EC.2. Proofs

Proof of Theorem 1. To facilitate the proof, we consider a variant of IPFC which works as follows: During period t, we fulfill item $i \in q$ from region j from facility k with probability $\mathbf{u}^{qt}_{kij} = \mathbf{U}^q_{kij}/(T\lambda^q_j)$ *regardless* of availability (i.e., only fulfill from facility 1 if it is the sampled facility). We denote this heuristic by MPFC, where the "M" stands for Modified. Since MPFC ignores inventory constraints, at the end of selling horizon, MPFC incurs a large penalty $c_p := \sum_j \sum_q \sum_k \left[\sum_{i \in q} c_{kij} + b_{kj}\right]$ for each violation of inventory constraints. Let $D^{qt}_j = 1$ if an order type q arrives from region j during period t and 0 otherwise. Total costs under MPFC is given by

$$C_{MPFC}(\theta) = \sum_{t,j,q,k} D^{qt}_j \left[\sum_{i \in q} c_{kij}\, X^{qt}_{kij}(\theta) \;+\; b_{kj} \max_{i \in q} \left\{X^{qt}_{kij}(\theta)\right\}\right]$$

$$+\, c_p \sum_{k,i} \left[\sum_{t,j,q\in i} D_j^{qt} X_{kij}^{qt}(\theta) - S_{ki}(\theta) \right]^+ ,$$

where $X_{kij}^{qt}(\theta)$ is a Bernoulli($\mathbf{u}_{kij}^{qt}$) random variable. Since c_p is larger than the cost of fulfilling an item from facility 1, $C_{IPFC}(\theta) \le C_{MPFC}(\theta)$. This allows us to bound $\mathbf{E}[C_{IPFC}(\theta)]/J^*(\theta)$ with $\mathbf{E}[C_{MPFC}(\theta)]/J^*(\theta)$. Moreover, since $J^*(\theta) \ge J_{DLP}(\theta)$, we can further bound $\mathbf{E}[C_{MPFC}(\theta)]/J^*(\theta)$ with $\mathbf{E}[C_{MPFC}(\theta)]/J_{DLP}(\theta)$. Note that the term with $[.]^+$ in $C_{MPFC}(\theta)$ is $O(\sqrt{\theta})$. To be precise, there exists a constant $M' > 0$ independent of $\theta > 0$, k, and i such that

$$\begin{aligned}
\mathbf{E}\left[\left(\sum_{t,j,q\in i} D_j^{qt} X_{kij}^{qt}(\theta) - S_{ki}(\theta)\right)^+\right] &\le \mathbf{E}\left[\left(\sum_{t,j,q\in i} (D_j^{qt} X_{kij}^{qt}(\theta) - \lambda_j^q \mathbf{u}_{kij}^{qt}(\theta))\right)^+\right] \\
&\quad + \mathbf{E}\left[\left(\sum_{t,j,q\in i} \lambda_j^q \mathbf{u}_{kij}^{qt}(\theta) - S_{ki}(\theta)\right)^+\right] \\
&\le \sum_{t,j,q\in i} \mathbf{E}\left[(D_j^{qt} X_{kij}^{qt}(\theta) - \lambda_j^q \mathbf{u}_{kij}^{qt}(\theta))^+\right] \\
&\le \sum_{t,j,q\in i} \lambda_j^q \sqrt{\mathrm{VAR}(D_j^{qt} X_{kij}^{qt}(\theta))} \;\le\; M'\sqrt{\theta},
\end{aligned}$$

where the second inequality follows because $\sum_{t,j,q\in i} \lambda_j^q \mathbf{u}_{kij}^{qt}(\theta) \le S_{ki}(\theta)$, the third inequality follows because $\mathbf{E}[D_j^{qt} X_{kij}^{qt}(\theta)] = \lambda_j^q \mathbf{u}_{kij}^{qt}(\theta)$, and the last inequality follows because $D_j^{qt} X_{kij}^{qt}(\theta)$ is a bounded random variable (hence, its variance must be bounded) and $T(\theta) = \theta T$.

Since $J_{DLP}(\theta) = \theta\, J_{DLP}$, the $O(\sqrt{\theta})$ term vanishes asymptotically as $\theta \to \infty$. So, we can focus on $\sum_{t,j,q,k} D_j^{qt} \left[\sum_{i\in q} c_{kij}\, X_{kij}^{qt}(\theta) + b_{kj} \max_{i\in q} \left\{X_{kij}^{qt}(\theta)\right\}\right]$. Observe that

$$\mathbf{E}\left[\max_{i\in q} \left\{X_{kij}^{qt}(\theta)\right\}\right] \;=\; \mathbf{E}\left[1 - \prod_{i\in q}(1 - X_{kij}^{qt}(\theta))\right] \;=\; 1 - \prod_{i\in q} \mathbf{E}\left[1 - X_{kij}^{qt}(\theta)\right].$$

Since $\mathbf{E}[X_{kij}^{qt}(\theta)] = \mathbf{u}_{kij}^{qt}$, $\mathbf{E}\left[\max_{i\in q} \left\{X_{kij}^{qt}(\theta)\right\}\right] \le \sum_{i\in q} \mathbf{u}_{kij}^{qt} \le |q| \max_{i\in q} \mathbf{u}_{kij}^{qt}$. So, we can bound:

$$\mathbf{E}\left[D_j^{qt}\left(\sum_{i\in q} c_{kij}\, X_{kij}^{qt}(\theta) + b_{kj} \max_{i\in q} \left\{X_{kij}^{qt}(\theta)\right\}\right)\right] \;\le\; \lambda_j^q \left(\sum_{i\in q} c_{kij} \mathbf{u}_{kij}^{qt} + b_{kj}|q| \max_{i\in q} \left\{\mathbf{u}_{kij}^{qt}\right\}\right).$$

Using the fact that $\mathbf{u}_{kij}^{qt} = \mathbf{u}_{kij}^{q1}$ for all $t \ge 1$ and the inequality $\frac{\sum_i (a_i + b_i c_i)}{\sum_i (a_i + b_i)} \le \frac{\sum_i b_i c_i}{\sum_i b_i}$ for all $a_i > 0$, $b_i > 0$, and $c_i \ge 1$, we have:

$$\lim_{\theta\to\infty} \frac{\mathbf{E}[C_{MPFC}(\theta)]}{J_{DLP}(\theta)} \le \frac{\sum_{q,j,k} \lambda_j^q \left(\sum_{i\in q} c_{kij} \mathbf{u}_{kij}^{qt} + b_{kj}|q| \max_{i\in q} \left\{\mathbf{u}_{kij}^{qt}\right\}\right)}{\sum_{q,j,k} \lambda_j^q \left(\sum_{i\in q} c_{kij} \mathbf{u}_{kij}^{qt} + b_{kj} \max_{i\in q} \left\{\mathbf{u}_{kij}^{qt}\right\}\right)}$$

$$\leq \frac{\sum_{q,j,k} \lambda_j^q b_{kj} |q| \max_{i\in q} \left\{\mathbf{u}_{kij}^{qt}\right\}}{\sum_{q,j,k} \lambda_j^q b_{kj} \max_{i\in q} \left\{\mathbf{u}_{kij}^{qt}\right\}} = \sum_{q,k,j} |q|\, F(q,k,j).$$

This completes the proof of Theorem 1. ■

Proof of Lemma 1. Since $\{\mathbf{y}_{kj}^{q1}\}$ is an optimal solution to J_{DLP}, it must satisfy $\mathbf{y}_{kj}^{q1} = \max_{i\in q} \left\{\mathbf{u}_{kij}^{q1}\right\}$. We will now provide a lower and an upper bound for the sum $\sum_k \max_{i\in q} \left\{\mathbf{u}_{kij}^{q1}\right\}$. The lower bound is straightforward:

$$\sum_k \max_{i\in q} \left\{\mathbf{u}_{kij}^{q1}\right\} \;\geq\; \sum_k \frac{\sum_{i\in q} \mathbf{u}_{kij}^{q1}}{|q|} \;=\; \frac{1}{|q|} \sum_{i\in q} \sum_k \mathbf{u}_{kij}^{q1} \;=\; 1,$$

where the last equality follows because $\sum_k \mathbf{u}_{kij}^{q1} = 1$. We now give an upper bound. Obviously, since $\mathbf{u}_{kij}^{q1} \leq 1$, we must have $\sum_k \max_{i\in q} \left\{\mathbf{u}_{kij}^{q1}\right\} \leq |S_K|$. But, also,

$$\sum_k \max_{i\in q} \left\{\mathbf{u}_{kij}^{q1}\right\} \;\leq\; \sum_k \sum_{i\in q} \mathbf{u}_{kij}^{q1} \;=\; \sum_{i\in q} \sum_k \mathbf{u}_{kij}^{q1} \;=\; |q|.$$

We conclude that $\sum_k \max_{i\in q} \left\{\mathbf{u}_{kij}^{q1}\right\} \leq \min\{|S_K|, |q|\}$. Since $F(q,k,j) = \lambda_j^q \mathbf{y}_{kj}^{q1} / \sum_{j',q',k'} \lambda_{j'}^{q'} \mathbf{y}_{k'j'}^{q'1}$, applying the above lower and upper bounds to $F(q,k,j)$ immediately yields the result. This completes the proof. ■

Proof of Theorem 2. The proof uses similar arguments as the proof og Theorem 1. The key is to show that $\mathbf{E}\left[\max_{i\in q} \{X_{kij}^{qt}\}\right] \leq B(|q|) \max_{i\in q} \{\mathbf{u}_{kij}^{qt}\}$ for all q,k,j. Fix (q,j) and let $|q| = n$ (whenever possible, we will suppress notational dependency on (t,q,j)). By our construction, for each k,

$$\mathbf{E}\left[\max_{i\in q} \{X_{ki}\}\right] \;\leq\; \sum_{m=1}^{n} \left[\frac{m}{n} M_k^m \;+\; m\left(1 - \frac{m}{n}\right) M_k^m\right],$$

where the expectation is taken with respect to the induced joint distribution. The first term in the summation follows because $\mathcal{I}_i^{km} = \mathcal{I}_{i'}^{km}$ for all i, $i' \in q$ with $\tilde{u}_{ki}^m = \tilde{u}_{ki'}^m = M_k^m > 0$, and $|\mathcal{I}_i^{km}| = \frac{m}{n} M_k^m$. The last term follows because, under the worst case scenario, the intervals $\mathcal{I}_i^{Akm}$ and $\mathcal{I}_{i'}^{Akm}$ may not intersect at all. Since there can be at most m such intervals (by definition, $\tilde{u}_k^m$ contains at most m non-zero elements), we have a multiplicative factor m. We divide our analysis into two cases. If n

is even, then $\max_{1\leq m\leq n}\ \left[\frac{m}{n}+m\left(1-\frac{m}{n}\right)\right]=\frac{n+2}{4}$. If n is odd, $\max_{1\leq m\leq n}\ \left[\frac{m}{n}+m\left(1-\frac{m}{n}\right)\right]=\frac{(n+1)^2}{4n}$. So, by definition of $B(.)$, $\mathbf{E}\left[\max_{i\in q}\ \{X_{ki}\}\right]\leq B(n)\sum_m M_k^m=B(|q|)\sum_m M_k^m$. But, $\max_{i\in q}\ \{\mathbf{u}_{ki}\}=\max_{i\in q}\{\sum_m \tilde{u}_{ki}^m\}=\sum_m \max_{i\in q}\{\tilde{u}_{ki}^m\}=\sum_m M_k^m$, where the second equality follows because, by construction, $M_k^m>0$ and $\tilde{u}_{ki}^m=0$ imply $\tilde{u}_{ki}^{m'}=0$ for all $m'<m$. We conclude that $\mathbf{E}\left[\max_{i\in q}\ \{X_{ki}\}\right]\leq B(|q|)\max_{i\in q}\{\mathbf{u}_{ki}\}$. The theorem now follows by the same argument as in the proof of Theorem 1.

■

EC.3. Numerical Study Description

In this appendix, we provide a detailed description of our numerical study in Section 5. Our numerical study was performed entirely using publicly-available data, and we provide enough detail here to allow readers to completely replicate our study. We first describe our numerical study domain (locations and distances), followed by the initial inventory placement and the actual simulation details.

Geographical Domain. Our study is placed in the continental United States. For customer locations, we start with the 100 largest metropolitan statistical areas (MSAs) as estimated by the US Census Bureau (2014), excluding Honolulu. We take into account the population of the MSAs in generating demand, so a more populous city generates proportionately more demand. Given the number of customer locations $|\mathcal{S}_J|$ that we are interested in, we simply select uniformly at random from this set of 99 cities.

For the list of potential facility locations, we use Chicago Consulting (2013), which reports the locations of the best $|\mathcal{S}_K|$ facilities for minimizing shipping cost in the US, for $|\mathcal{S}_K|=1,2,\ldots,10$. We remove the Puerto Rico locations from this list and select networks with $|\mathcal{S}_K|=2,5$, and 9. Thus, although we do not optimize the location of the facility ourselves (optimal location of fulfillment centers for ecommerce is a different research question), our chosen fulfillment facility locations arguably are somewhat close to optimal.

We use UPS ground shipping rates to estimate our shipping cost function. With 99 destination cities and 9 potential facility locations, there are 891 possible origin-destination pairs. For each such pair, we get the shipping rate from UPS for a package of weight 1, 2, or 3 pounds, choosing package weight uniformly at random. We then first estimated the following linear shipping cost model, where d_{kj} is the distance in miles from facility k to customer region j and $n_{kj} \in \{1,2,3\}$ is the number of items shipped, assuming each item weighs exactly one pound:

$$cost(n_{kj},k,j) \;=\; \beta_0 + \beta_1 n_{kj} + \beta_2 d_{kj} + \beta_3 n_{kj} d_{kj}$$

The estimate of coefficient β_2 was insignificant in the above estimation, so we removed it and re-estimated the parameters. This resulted in the following final estimate, with an R^2 of 94.5% and p-values of the order of 10^{-15}:

$$cost(n_{kj},k,j) \;=\; 8.759 + 0.423 n_{kj} + 0.000541 n_{kj} d_{kj}$$

In our entire numerical study, we either used the shipping cost function given above or minimized the number of total shipments. However, our methodology is such that any shipping cost function should be easily usable. Note also that we use a fictitious facility indexed $k=1$ with infinite supply to model the costs incurred due to stockouts. To make the problem reasonable, facility 1 should have higher costs than regular facilities. We implement this using a penalty factor of 2. That is, we set $b_{1j} = 2 \times 8.759$ and $c_{1ij} = 2 \times (0.423 + 0.00541 \max_{k,j} d_{kj})$.

Demand Forecasts and Initial Inventory. First, we describe how we construct the demand rates λ in our simulation, given a set of items $\mathcal{S}_I = \{1,2,\ldots,|\mathcal{S}_I|\}$. The main research problem addressed in this paper is how to fulfill an order that contains more than one item. So, we need to generate demand rates for each order types. However, for both real-world and analytical tractability reasons, we cannot consider all possible order types in $2^{\mathcal{S}_I}$. So, we consider a smaller set of order types, defined by two parameters: n_{max}, which denotes the maximum number of items contained in an order (i.e., maximum order size), and n_0, which denotes the number of order types of each size less than or equal to n_{max}. For the most part, we use $n_{max} = n_0 = 5$, (adding the "empty" order $q = \emptyset$ gives $|\mathcal{S}_Q| = n_{max}.n_0 + 1 = 26$ as mentioned in Section 5) although we test various other values of both parameters. Given n_{max}, we first generate the total probability $p(n)$ of all orders of sizes $0, 1, 2, \ldots, n_{max}$. The probability $p(n)$ is chosen uniformly at random so that $p(n) \in [0,1]$ $\forall n$ and $\sum_0^{n_{max}} p(n) = 1$.

Let λ^q denote the total demand rate for order type q from all regions; that is, $\lambda^q = \sum_{j \in J} \lambda_j^q$. Given $p(n)$, we first generate λ^q as follows. For each $n \in \{1, 2, \ldots, n_{max}\}$, we select uniformly at random $\min\{n_0, \binom{|\mathcal{S}_I|}{n}\}$ subsets of $\mathcal{S}_I^n$ to have positive demand rates. Let $Q(n)$ denote this subset. We then choose $\{\lambda^q\}_{q \in Q(n)}$ to be uniformly at random in $[0, p(n)]^{|Q(n)|}$ such that $\sum_{q \in Q(n)} \lambda^q = p(n)$. Lastly, we generate λ_j^q by simply scaling λ^q to the population of each city $j \in \mathcal{S}_J$. That is, $\lambda_j^q = \lambda^q.pop(j) / \sum_{j \in \mathcal{S}_J} pop(j)$, where $pop(j)$ is the population of the metropolitan statistical area j.

We now describe the initial inventory placement. In practice, it is often the case that any given single facility stocks only a subset of all the items sold by the retailer; this may be because of supplier considerations, material handling requirements, equipment, capacity constraints, etc. To model this, we use a parameter $p_{stock} \in [0,1]$ such that, for any given facility k and item i, the probability that k stocks i is p_{stock}. That is, $P(S_{ki} > 0) = p_{stock}$, i.i.d. for all k, i. In our numerical studies we do test the sensitivity of our results with respect to p_{stock}, including the case $p_{stock} = 1$ where every facility stocks every item. For most of our experiments, we use $p_{stock} = 0.75$. Now, consider a facility k and item i for which we have determined that $S_{ki} > 0$. How should we set the initial inventory level S_{ki}? It is reasonable to believe that the firm would compute the expected demand from all customers for whom this facility is the nearest that stocks item i, and keep inventory level equal to the expected total demands plus some safety stock, calculated in a newsvendor fashion. This is exactly what we do. Formally, for a given facility k and item i, we first find the set of customers (or regions) $J(k,i)$ for whom it should stock the inventory: $J(k,i) = \{j \in \mathcal{S}_J : d_{kj} = \min_{k' \in K : S_{k'i} > 0} d_{k'j}\}$. Define $\lambda(k,i) = \sum_{j \in J(k,i), q \ni i} \lambda_j^q$; this is the total incoming expected demand to facility k for item i from all orders that contain i. We then compute the expected value and standard deviation of the demand for item i from $J(k,i)$, given the time horizon θT, as follows: $\mu(k,i) = \theta T \lambda(k,i)$ and $\sigma(k,i) = \sqrt{\theta T \lambda(k,i)(1 - \lambda(k,i))}$. Next, given another global parameter CSL (Cycle Service Level), we simply use the newsvendor fractile at that level to determine the starting inventory: $S_{ki} =$

$\mu(k,i) + z_{CSL}\sigma(k,i)$, where z_{CSL} is the inverse of the standard normal distribution at probability CSL. Our default value for CSL in the numerical experiments is 0.5.

Simulation Procedure. Given our setup above, our simulation process is fairly simple. Once the parameters $|\mathcal{S}_I|, |\mathcal{S}_J|, |\mathcal{S}_K|, n_{max}, n_0, p_{stock}, CSL$, and θT are defined, we generate the sets $\mathcal{S}_I, \mathcal{S}_J, \mathcal{S}_K$, and $\mathcal{S}_Q$, and the matrices λ, c, b, and S. We then compute the values $\{u_{kij}^{qt}\}$ that define the IPFC algorithm, as well as the values $\{g_j^q\}$ which defines the DPFC algorithm. We also implement a myopic algorithm, which works as follows: given an order q at time t from customer region j, it simply fulfills every item in q from the facility nearest to j that has positive inventory of that item.

We then generate a single demand sequence based on λ and θT. All three algorithms are applied to *the same demand sequence*. Therefore, the variation in the demand affects all three algorithms equally, and this allows for a better comparison of the algorithms. This constitutes one simulation trial. For each setting of the control parameters such as $|\mathcal{S}_I|, |\mathcal{S}_J|, p_{stock}, n_{max}$ etc., we run 30 simulation trials, where each simulation trial has a new demand rate λ and therefore a new demand sequence, generated as defined above. This allows us to obtain statistical significance in our results, as detailed in Section 5. In total, we ran over a few thousand simulation trials with several different combinations of parameters. A selection of these that are particularly insightful are reported in Section 5. In our computing environment, the time taken for a typical single simulation trial is about 2 to 3 minutes.

References

Chicago Consulting. 2013. 10 best warehouse networks 2013. http://www.chicago-consulting.com/10best.shtml.

US Census Bureau. 2014. Annual estimates of the resident population: April 1, 2010 to July 1, 2013 - united states - metropolitan and micropolitan statistical area.