Manufacturing & Service Operations Management

Sequencing Appointments for Service Systems Using Inventory Approximations

Ho-Yin Mak
Department of Industrial Engineering and Logistics Management
The Hong Kong University of Science and Technology, Kowloon, Hong Kong
hymak@ust.hk

Ying Rong
Antai College of Economics and Management
Shanghai Jiao Tong University, Shanghai 200052, China
yrong@sjtu.edu.cn

Jiawei Zhang
Stern School of Business, IOMS-Operations Management
New York University, New York, NY 10012
jzhang@stern.nyu.edu

Online Appendix A: Proofs of Analytical Results

Proof of Proposition 1

Since we consider the minimization problem in (7), relaxing (8) leads to a lower bound for $C^*(\mathbf{x})$ for any $\mathbf{h} \in \mathbf{H}$. Thus, maximizing over $\mathbf{h}$ in $\mathbf{H}$ also provides a lower bound. $\square$

Proof of Proposition 2

First, we note that

$$\min_{\mathbf{x}\in\mathbf{X}} \max_{\mathbf{h}\in\mathbf{H}} \min_{\mathbf{S}} \sum_{i=1}^{N-1} \tilde{c}_i^{SB}(\mathbf{S}_i, \mathbf{h}, \mathbf{x}) = \min_{\mathbf{x}\in\mathbf{X}} \min_{\mathbf{S}} \max_{\mathbf{h}\in\mathbf{H}} \sum_{i=1}^{N-1} \tilde{c}_i^{SB}(\mathbf{S}_i, \mathbf{h}, \mathbf{x}).$$

The equality holds because $\sum_{i=1}^{N-1} \tilde{c}_i^{SB}(\mathbf{S}_i, \mathbf{h}, \mathbf{x})$ is convex in $\mathbf{S}$ and concave in $\mathbf{h}$, given any fixed $\mathbf{x} \in \mathbf{X}$. The minimization over $\mathbf{S}$ is over a convex set, and the maximization over $\mathbf{h}$ is over a compact, convex set. Applying the saddle point theorem (e.g., Aubin and Ekeland [2], pg. 295), the ordering of maximization and minimization can be interchanged.

After the interchange, the inner problem $\max_{\mathbf{h}\in\mathbf{H}} \sum_{i=1}^{N-1} \tilde{c}_i^{SB}(\mathbf{S}_i, \mathbf{h}, \mathbf{x})$ is a linear program. By introducing dual variables $\mathbf{y}$ and $\mathbf{z}$ associated with constraints in the definition of $\mathbf{H}$, strong duality implies that $\max_{\mathbf{h}\in\mathbf{H}} \sum_{i=1}^{N-1} \tilde{c}_i^{SB}(\mathbf{S}_i, \mathbf{h}, \mathbf{x})$ is equivalent to

$$\min_{\mathbf{y},\mathbf{z}} \sum_{j=1}^{N-1} \kappa y_j + \sum_{i=1}^{N-1} \sum_{m=1}^{N} p_m x_{i+1,m} \bar{B}_i^{SB}(\mathbf{S}_i, \mathbf{x}) \tag{A.1}$$

$$\begin{aligned}
s.t. \quad & y_1 \geq \bar{I}_1^{SB}(\mathbf{S}_i, \mathbf{x}) & & \\
& y_1 + z_{i1} \geq \alpha_{i1} \bar{I}_i^{SB}(\mathbf{S}_i, \mathbf{x}), & & i = 2, ..., N-1 \\
& y_i - z_{i,i-1} \geq \alpha_{ii} \bar{I}_i^{SB}(\mathbf{S}_i, \mathbf{x}), & & i = 2, ..., N-1 \\
& y_j + z_{ij} - z_{i,j-1} \geq \alpha_{ij} \bar{I}_i^{SB}(\mathbf{S}_i, \mathbf{x}), & & j = 2, ..., i-1; \quad i = 2, ..., N-1 \\
& z_{ij} \geq 0, & & j = 1, ..., i-1; \quad i = 2, ..., N-1.
\end{aligned} \tag{A.2}$$

Therefore, we have shown that $\min_{\mathbf{x} \in \mathbf{X}} \min_{\mathbf{S}} \max_{\mathbf{h} \in \mathbf{H}} \sum_{i=1}^{N-1} \tilde{c}_i^{SB}(\mathbf{S}_i, \mathbf{h}, \mathbf{x})$ is equivalent to

$$\begin{aligned}
\min_{\mathbf{x} \in \mathbf{X}} \min_{\mathbf{S}} \min_{\mathbf{y}, \mathbf{z}} \quad & \sum_{j=1}^{N-1} \kappa y_j + \sum_{i=1}^{N-1} \sum_{m=1}^{N} p_m x_{i+1,m} \bar{B}_i^{SB}(\mathbf{S}, \mathbf{x}) \\
s.t. \quad & \text{(A.2)}.
\end{aligned} \tag{A.3}$$

Now denote $a_i = \bar{I}_i^{SB}(\mathbf{S}_i, \mathbf{x})$ and $c_i = \sum_{j=1}^{i} S_{ij} - \sum_{j=1}^{i} \sum_{m=1}^{N} \mu_m x_{jm}$. Based on the definitions of $\bar{I}_i^{SB}(\mathbf{S}_i, \mathbf{x})$ and $\bar{B}_i^{SB}(\mathbf{S}_i, \mathbf{x})$ in (17) and (18), we have $\bar{B}_i^{SB}(\mathbf{S}_i, \mathbf{x}) = a_i - c_i$. Then, (A.3) becomes

$$\min_{\mathbf{x}, \mathbf{y}, \mathbf{z}, \mathbf{s}, \mathbf{a}, \mathbf{c}} \sum_{j=1}^{N-1} \kappa y_j + \sum_{i=1}^{N-1} \sum_{m=1}^{N} p_m x_{i+1,m} (a_i - c_i) \tag{A.4}$$

$$\begin{aligned}
s.t. \quad & y_1 \geq a_1 & & \\
& y_1 + z_{i1} \alpha_{i1} a_i, & & i = 2, ..., N-1 \\
& y_i - z_{i,i-1} \geq \alpha_{ii} a_i, & & i = 2, ..., N-1 \\
& y_j + z_{ij} - z_{i,j-1} \geq \alpha_{ij} a_i, & & j = 2, ..., i-1; i = 2, ..., N-1 \\
& z_{ij} \geq 0, & & j = 1, ..., i-1; i = 2, ..., N-1
\end{aligned}$$

$$a_i = \frac{1}{2} c_i + \frac{1}{2} \sqrt{c_i^2 + \sum_{j=1}^{i} \sum_{m=1}^{N} \sigma_m^2 x_{jm}^2}, \qquad i = 1, ..., N-1 \tag{A.5}$$

$$c_i = \sum_{j=1}^{i} S_{ij} - \sum_{j=1}^{i} \sum_{m=1}^{N} \mu_m x_{jm}, \qquad i = 1, ..., N-1. \tag{A.6}$$

We define the decision variable $r_{i+1,m} = x_{i+1,m}(a_i - c_i)$. This can be imposed by adding constraints (22) and $r_{im} \geq 0$, with sufficiently large U_{im} value (e.g., [1]). Next, S_{ij} only appears in constraints (A.6). Therefore, this set of constraints are redundant and can be removed. In addition, the equality sign in (A.5) can be replaced with the $\geq$ sign, because the objective value is increasing in a_i and decreasing in c_i. By making those transformations, we obtain (21). $\square$

Proof of Proposition 3

To prove Proposition 3, we first present the following result by utilizing Scarf's newsvendor bound. Define $\sigma_j(\mathbf{x}) = \sum_{i=1}^{N} \sigma_m x_{jm}$. By the definition of $\tilde{c}_i^{SB}(\mathbf{S}_i, \mathbf{h}, \mathbf{x})$, we have

$$\min_{\mathbf{S}_i} \tilde{c}_i^{SB}(\mathbf{S}_i, \mathbf{h}, \mathbf{x}) = \sqrt{p_{i+1}(\mathbf{x}) \left(\sum_{j=1}^{i} (\sigma_j(\mathbf{x}))^2 \right) \left(\sum_{j=1}^{i} \alpha_{ij} h_{ij} \right)},$$

and thus problem (20) is equivalent to $\min_{\mathbf{x}\in\mathbf{X}} \max_{\mathbf{h}\in\mathbf{H}} \tilde{C}^{SB}(\mathbf{h},\mathbf{x})$ where

$$\tilde{C}^{SB}(\mathbf{h},\mathbf{x}) = \sum_{i=1}^{N-1} \sqrt{p_{i+1}(\mathbf{x}) \left(\sum_{j=1}^{i} (\sigma_j(\mathbf{x}))^2\right) \left(\sum_{j=1}^{i} \alpha_{ij} h_{ij}\right)}. \tag{A.7}$$

Consider some feasible sequence denoted by $\mathbf{x}^A$, in which jobs m_1 and m_2 are assigned the j-th and k-th positions, with $j < k$, respectively, and let $\mathbf{h}^A \in \mathbf{H}$ be an optimal solution to $\max_{\mathbf{h}\in\mathbf{H}} \tilde{C}^{SB}(\mathbf{h},\mathbf{x}^A)$. Then, consider the sequence obtained by interchanging the positions of jobs m_1 and m_2 under $\mathbf{x}^A$, and denote the new sequence by $\mathbf{x}^B$. Let $\mathbf{h}^B$ be an optimal solution to $\max_{\mathbf{h}\in\mathbf{H}} \tilde{C}^{SB}(\mathbf{h},\mathbf{x}^B)$. Then, by applying (A.7),

$$\begin{aligned}
\tilde{C}^{SB}(\mathbf{x}^A,\mathbf{h}^A) &= \sum_{i=1}^{N-1} \sqrt{p_{i+1}(\mathbf{x}^A) \left(\sum_{j=1}^{i} (\sigma_j(\mathbf{x}^A))^2\right) \left(\sum_{j=1}^{i} \alpha_{ij} h_{ij}^A\right)} \\
&\leq \sum_{i=1}^{N-1} \sqrt{p_{i+1}(\mathbf{x}^B) \left(\sum_{j=1}^{i} (\sigma_j(\mathbf{x}^A))^2\right) \left(\sum_{j=1}^{i} \alpha_{ij} h_{ij}^A\right)} \\
&\leq \sum_{i=1}^{N-1} \sqrt{p_{i+1}(\mathbf{x}^B) \left(\sum_{j=1}^{i} (\sigma_j(\mathbf{x}^B))^2\right) \left(\sum_{j=1}^{i} \alpha_{ij} h_{ij}^A\right)} \\
&= \tilde{C}^{SB}(\mathbf{x}^B,\mathbf{h}^A) \\
&\leq \max_{\mathbf{h}\in\mathbf{H}} \tilde{C}^{SB}(\mathbf{x}^B,\mathbf{h}) = \tilde{C}^{SB}(\mathbf{x}^B,\mathbf{h}^B).
\end{aligned}$$

In this set of inequalities, we incorporate the impact of interchange on the late-start penalty, variance and local holding cost allocation stepwise. Note that $p_j(\mathbf{x}^A) = p_k(\mathbf{x}^B) = p_{m_1} \geq p_j(\mathbf{x}^B) = p_k(\mathbf{x}^A) = p_{m_2}$, and $\sqrt{\sum_{j=1}^{i} (\sigma_i(\mathbf{x}^A))^2}$ is strictly increasing in i. Then, the first inequality above holds due to the inequality:$a_1a_2 + b_1b_2 \geq a_1b_2 + b_1a_2$ for $a_i \geq b_i$, $i = 1,2$. The second inequality holds because $\sigma_i(\mathbf{x}^A) = \sigma_i(\mathbf{x}^B)$ for $i < j$ and $i \geq k$, and $\sigma_i(\mathbf{x}^A) < \sigma_i(\mathbf{x}^B)$ for $i = j, ..., k-1$. Therefore, interchanging the positions of m_1 and m_2, and re-optimizing $\mathbf{h}$ (in the last inequality) leads to an increase in cost. This completes the proof. □

Online Appendix B: The Connection between Serial Supply Chain Inventory Problems and Appointment Scheduling Problems

In this Appendix, we demonstrate how the formulation of the decomposed subproblem $\min_{\mathbf{S}_i} \tilde{c}_i(\mathbf{S}_i,\mathbf{h},\mathbf{x})$ in Proposition 1, for each particular i, can be interpreted as the local base stock optimization for a serial supply chain inventory problem.

We adopt the inventory model notation from Zipkin [13, Section 8.3] for the continuous review setting. We note that both continuous and periodic review will give rise to notationally equivalent models, but we utilize the former for the ease of explanation. In particular, there are i stages in the serial supply chain, with the most upstream denoted by $j = 1$ and the most downstream denoted by $j = i$. The most downstream stage faces stochastic demand from an external source, and all other stages face orders from their respective immediate downstream neighbors. The most upstream stage is replenished by an external supplier with ample stock, and all other stages are fed by their respective immediate upstream neighbors. Replenishment orders by stage j (to the external supplier if $j=1$ and to stage $j-1$ otherwise) are subject to a deterministic lead time of L_j. Downstream orders are met with stock on hand whenever possible, and are backordered otherwise. Note that while backorders may exist at all stages, penalty costs are charged only for backorders at the most downstream stage facing external demand. It has been shown by many researchers, including Clark and Scarf [5], Chen and Zheng [4], Chen and Song [3] and Huh and Janakiraman [9], that a base stock policy is optimal under this setting, as well as a variety of other settings, such as the case with periodic review. In general, Huh and Janakiraman [9] show that a base stock policy is optimal as long as the evolution of the demand process is not affected by the ordering policy.

Following a base stock policy, whenever the immediate downstream stage $j+1$ places an order, stage j places an order to raise its inventory position (i.e., net inventory plus inventory on order from stage $j-1$) back to the pre-determined base stock level. The base stock levels are optimized to balance underage and overage costs. We assume that there are no restrictions (e.g., capacity constraints) on order quantities. Note that, in the presence of general capacity constraints, the system may not be stable (i.e., no unique stationary distribution exists, see, e.g., [10]). Glasserman and Tayur [8] and Huh *et al.* [10] prove that, a unique stationary distribution indeed exists as long as the ordering capacity at each stage exceeds the expected demand, which holds for our uncapacitated system in which ordering capacities are infinite for all stages.

Because problem (11) involves $\sum_{i=1}^{N-1} \tilde{c}_i(\mathbf{S}_i, \mathbf{h}, \mathbf{x})$, we draw a connection to a collection of $N-1$ serial supply chain inventory problems, indexed by $i=1,...,N-1$. To resemble the job sequencing problem, we distinguish between "nodes" and "stages" in the serial supply chain. We denote a node as a production facility and denote stage as the position of the production facility in the serial supply chain. That is, a node corresponds to a job and a stage corresponds to a position in a job sequence. We allow the shuffling of nodes among stages of the supply chains. Moreover, the lead time L_m can be treated as the production time at node m so that L_m is associated with node m instead of the stage. We first proceed with the following notation:

<u>Demand and Cost Parameters</u>

- $d(t_1,t_2)$: random external demand during the time interval t_1 to t_2 (where $t_1 < t_2$);
- $L_j(\mathbf{x})$: deterministic replenishment lead time of stage j when ordering from stage $j-1$ for a given sequence $\mathbf{x}$;
- h_{ij} : the local holding cost per unit at stage j of serial supply chain i;
- $p_{i+1}(\mathbf{x})$: the shortage cost per unit at the most downstream stage of supply chain i for a given sequence $\mathbf{x}$.

<u>Decision Variables</u>

- S_{ij} : the local (installation) base stock level at stage j in supply chain i;
- $\mathbf{S}_i$: the vector $(S_{i1}, S_{i2}, ..., S_{ii})$ of local base stock levels in supply chain i.

<u>Performance Indicators</u>

Given a sequence $\mathbf{x}$ and base stock level vector $\mathbf{S}_i$, we consider the following at time t:

- $B_{ij}(\mathbf{S}_i, \mathbf{x}, t)$: the amount of backorders at stage j of supply chain i;
- $I_{ij}(\mathbf{S}_i, \mathbf{x}, t)$: the on-hand inventory at stage j of supply chain i.

Referring to Zipkin [13, Section 8.3], the inventory dynamics for stage j of supply chain i can be expressed as follows:

$$I_{ij}\left(\mathbf{S}_i, \mathbf{x}, t + \sum_{k=1}^{j} L_k(\mathbf{x})\right) = \left[S_{ij} - B_{i,j-1}\left(\mathbf{S}_i, \mathbf{x}, t + \sum_{k=1}^{j-1} L_k(\mathbf{x})\right), d\left(t + \sum_{k=1}^{j-1} L_k(\mathbf{x}), t + \sum_{k=1}^{j} L_k(\mathbf{x})\right)\right]^{+}$$

$$B_{ij}\left(\mathbf{S}_i, \mathbf{x}, t + \sum_{k=1}^{j} L_k(\mathbf{x})\right) = \left[S_{ij} - B_{i,j-1}\left(\mathbf{S}_i, \mathbf{x}, t + \sum_{k=1}^{j-1} L_k(\mathbf{x})\right), d\left(t + \sum_{k=1}^{j-1} L_k(\mathbf{x}), t + \sum_{k=1}^{j} L_k(\mathbf{x})\right)\right]^{-}$$

Because the intervals $\left[t + \sum_{k=1}^{j-1} L_k(\mathbf{x}), t + \sum_{k=1}^{j} L_k(\mathbf{x})\right)$ for different j are disjoint and the demand process is time-independent, $I_{ij}(\mathbf{S}_i, \mathbf{x}, t)$ and $B_{ij}(\mathbf{S}_i, \mathbf{x}, t)$ are time-stationary (in distribution). Therefore, with a slight abuse of notation, we drop the time index. Then, we denote $d_j(\mathbf{x}) = d\left(t + \sum_{k=1}^{j-1} L_k(\mathbf{x}), t + \sum_{k=1}^{j} L_k(\mathbf{x})\right)$ as the random demand during the replenishment lead time of stage j. Therefore, with the above interpretation of our notation, one may verify that the inventory dynamics of supply chain i can also be written as (9) and (10) under the steady state.

Thus, the total expected holding and backorder cost for supply chain i, as given by (Zipkin [13, Equation 8.3.5]), is exactly of the same form as $\tilde{c}_i(\mathbf{S}_i, \mathbf{h}, \mathbf{x})$; i.e.,

$$\tilde{c}_i(\mathbf{S}_i, \mathbf{h}, \mathbf{x}) = \sum_{j=1}^{i} h_{ij} E\left[I_{ij}(\mathbf{S}_i, \mathbf{x})\right] + p_{i+1}(\mathbf{x}) E\left[B_{ii}(\mathbf{S}_i, \mathbf{x})\right]. \tag{A.8}$$

With the independence of demand quantities for different j, given the values of $\mathbf{x}$, one can solve (A.8) using a convex decomposition method recursively, with only one decision variable involved in each step. However, though such an optimization approach can potentially be applied to the decomposed supply chains individually, such an implementation is not sufficient to evaluate the approximate appointment scheduling cost formulation (11) that requires maximization over the holding cost allocation for a given sequence variable $\mathbf{x}$. Moreover, our ultimate goal is to solve the combinatorial job sequencing problem by optimizing over $\mathbf{x}$, which requires the use of an integer programming solvers. To formulate the problem in a form that can be handled by such solvers, it helps to obtain closed-form approximations of the serial supply chain costs, instead of using the algorithmic procedures to compute them. Shang and Song [12], Gallego and Ozer [6], among others, have proposed simple newsvendor-like heuristics to approximate the base stock levels and cost function. In particular, the approximations by Shang and Song [12] can be applied to not only the continuous review setting, but also the periodic review setting. Moreover, their heuristic does not require any specific distributional form of $d_j(\mathbf{x})$. Therefore, we adopt this approximation in our solution procedure.

One subtle point to note is that, in the inventory model, the independent random variables $d_j(\cdot)$ arise from the increments over non-overlapping time intervals of the same stochastic process of demand (e.g., Poisson or compound Poisson process). However, in the appointment planning problem, the job durations $d_j(\cdot)$ can be completely unrelated. We note that the results from inventory theory that we apply do not make use of the fact that $d_j(\cdot)$ random variables arise from the a common stochastic process.

Online Appendix C: Mixed Integer Linear Programming Formulation for the Job Sequencing Problem

We provide a mixed integer linear programming formulation for the job sequencing problem using SAA. The main idea of SAA is to solve stochastic problems using a deterministic equivalent formulation obtained by sampling the duration distribution. We use this formulation as a benchmark for our computational studies in Section 4. This formulation can be solved using commercial solvers, such as CPLEX. The following notation will be needed:

Input Parameters

N : the number of jobs;

W : the sample size and w denotes the index of individual samples;

d_{mw} : realized duration of job m under sample w.

U : A large constant introduced for linearization of the objective function.

Decision Variables

$x_{jm} \in \{0,1\}$: binary variable where $x_{jm} = 1$ if job m is assigned the j-th position;

s_j : time allowance for the job in position j;

γ_{jw} : idle time for operating room before the job in position $j+1$ under sample w;

β_{jw} : waiting time for the job in position $j+1$ under sample w;

r_{jm} : the sample average waiting time for job m if it is assigned to position j, 0 otherwise.

Using SAA, suppose we draw a sufficient number (W) of samples. Then, the job sequencing problem can be formulated as shown below.

$$\min_{\boldsymbol{x},\boldsymbol{s},\boldsymbol{\gamma},\boldsymbol{\beta}} \frac{1}{W}\sum_{j=1}^{N-1}\sum_{w=1}^{W}\kappa\gamma_{jw} + \sum_{j=2}^{N}\sum_{m=1}^{N} p_m r_{jm} \tag{A.9}$$

$$\text{s.t.} \qquad s_1 + \beta_{1w} - \gamma_{1w} - \sum_{m=1}^{N} d_{mw}x_{1m} = 0, \quad w = 1, ..., W \tag{A.10}$$

$$s_j + \beta_{jw} - \gamma_{jw} - \beta_{j-1,w} - \sum_{m=1}^{N} d_{mw} x_{jm} = 0, \quad j = 2, ..., N-1 \text{ and } w = 1, ..., W \tag{A.11}$$

$$\sum_{j=1}^{N} x_{jm} = 1, \quad m = 1, ..., N \tag{A.12}$$

$$\sum_{m=1}^{N} x_{jm} = 1, \quad j = 1, ..., N \tag{A.13}$$

$$-r_{jm} + U x_{j,m} + \frac{1}{W}\sum_{w=1}^{W} \beta_{j-1,w} \le U, \quad j = 1, ..., N-1 \text{ and } m = 1, ..., N \tag{A.14}$$

$$\gamma_{jw}, \beta_{jw}, r_{jm} \ge 0 \tag{A.15}$$

$$x_{jm} \in \{0, 1\}. \tag{A.16}$$

(A.14) together with nonnegativity of r_{jm} implies $r_{jm} = \frac{1}{W}\sum_{w=1}^{W} \beta_{j-1,w}$ if $x_{j,m} = 1$, and 0 otherwise. In other words, r_{jm} represents the sample average waiting time for job m if it is assigned to position j. Therefore, the objective function (A.9) represents the sample average cost for job sequencing problem. Note that U should be an upper bound on the value of $\frac{1}{W}\sum_{w=1}^{W} \beta_{j-1,w}$.

Ge et al. [7] show that, under (A.10), (A.11) and nonnegativity of γ_{jw} and β_{jw}, there exists an optimal solution such that at least one of γ_{jw} and β_{jw} is zero. The intuition is that when the per unit idle time costs are uniform, it is never beneficial to keep the server idle while the next job is waiting to begin. (A.12) and (A.13) are the assignment constraints for jobs. Therefore, such a linear integer programming formulation is valid.

Finally, if we fix x_{jm}, we can evaluate the cost for job scheduling problem for a given sequence.

Online Appendix D: Enhancements to Improve Computational Speed for MISOCP Formulations

Proposition 3 allows us to tighten the MISOCP formulation by adding extra constraints that remove part of the feasible region, but not the optimal solution, via the following result.

COROLLARY 1. *If $\sigma_{m_1} < \sigma_{m_2}$ and $p_{m_1} > p_{m_2}$ for jobs m_1 and m_2, then the following inequality*

$$x_{i,m_1} + x_{j,m_2} \le 1, \quad i > j. \tag{A.17}$$

is a valid constraint to the MISOCP formulation (21).

In addition, we define three different sets corresponding to each job m.

1. $\Phi_m := \{i : \sigma_i < \sigma_m \text{ and } p_i > p_m\}$: Φ_m denotes the set of jobs that should be in front of job m in the optimal solution;

2. $\Xi_m := \{i : \sigma_i > \sigma_m \text{ and } p_i < p_m\}$: Ξ_m denotes the set of jobs that should be behind job m in the optimal solution;

3. $\varphi_m := \{1, ..., N\} \setminus (\{m\} \cup \Phi_m \cup \Xi_m)$: φ_m denotes the remaining set of jobs.

Let $|\Phi_m|$, $|\Xi_m|$ and $|\varphi_m|$ be the size of these three sets. Then we can fix some of x_{jm} variables to be zeros for some assignments that will be suboptimal by the following corollary. Qualitatively, it means that job m cannot be sequenced at any of the first $|\Phi_m|$ or last $|\Xi_m|$ positions.

COROLLARY 2. *For each m, we have $x_{jm} = 0$ for $j \le |\Phi_m|$ or $j \ge N - |\Xi_m| + 1$.*

The next result allows us to obtain tight U_{im} values for the linearization procedure. It is well known (e.g., Adams *et al.* [1]) that tighter values of such linearization constants produce tighter continuous relaxations of the associated integer program, and thus better computation speeds in general.

PROPOSITION 1. *Let job $\chi_m(k)$ have the k-th largest variance in set φ_m; i.e., $\sigma_{\chi_m(k)} \ge \sigma_{\chi_m(j)}$ for $k < j$. Then, for all $m = 1, ..., N$, the following are valid upper bound constants for linearization:*

$$U_{im} = \max_{\{k \neq m: ||\Phi_k| \le i-1, |\Xi_k| \le N-i\}} \sqrt{\kappa/p_k} \sqrt{\sum_{j=1}^{i-|\Phi_k|-1} \sigma^2_{\chi_k(j)} + \sum_{j\in\Phi_k} \sigma_j^2}.$$

Proof: Observe that in (22), we only need U_{im} to be an upper bound on the associated expected backorder value of chain $i-1$ when $x_{im} = 0$; i.e., job m is not assigned the $i-$th position. This can be obtained by computing the expected backorders for assigning job k to the i-th position, and taking the maximum over $k \neq m$. To do that, we may arrange the jobs with largest variances (excluding those eliminated by the partial ordering relationship) in the early positions, as we show below.

Let $\mathbf{x}^*, \mathbf{h}^*, \mathbf{S}^*$ be the optimal solution to (21) with $x_{im} = 0$ enforced. Let $\sigma_j(\mathbf{x}^*)$ and $p_j(\mathbf{x}^*)$ be the corresponding demand standard deviation and penalty cost for position j under optimal sequence. In addition, let $\bar{B}^{SB*}_{i-1}$ be the optimal value of $\bar{B}^{SB}_{i-1}$ in (21) where $\bar{B}^{SB}_{i-1} = a_{i-1} - c_{i-1}$ under the optimal solution. Then, we can write

$$\begin{aligned}
\bar{B}^{SB*}_{i-1} &= \sqrt{\frac{\sum_{j=1}^{i-1} h^*_{i-1,j}/(i-1)}{p_i(\mathbf{x}^*)}} \sqrt{\sum_{j=1}^{i-1} \sigma_j^2(\mathbf{x}^*)} \\
&\le \sqrt{\kappa/p_i(\mathbf{x}^*)} \sqrt{\sum_{j=1}^{i-1} \sigma_j^2(\mathbf{x}^*)} \\
&\le \max_{\{k \neq m: ||\Phi_k| \le i-1, |\Xi_k| \le N-i\}} \sqrt{\kappa/p_k} \sqrt{\sum_{j=1}^{i-|\Phi_k|-1} \sigma^2_{\chi_k(j)} + \sum_{j\in\Phi_k} \sigma_j^2}.
\end{aligned}$$

The first equality can be obtained directly by substituting the optimal solution of $\min_{\mathbf{S_i}} \tilde{c}_i^{SB}(\mathbf{S_i}, \mathbf{h}, \mathbf{x})$. The second inequality is due to the fact that $h_{ij} \le \kappa$ for $\mathbf{h} \in \mathbf{H}$. $\square$

Online Appendix E: A Lagrangian Duality-Based Derivation of Proposition 1

In this Appendix, we show that formulation (11) is equivalent to the Lagrangian dual of (7) obtained by relaxing a constraint equivalent to (8) that is provided in Lemma 1 below. The intuition of Lemma 1 is that, the expected idle time, $E[I_{ij}(\mathbf{S}_i, \mathbf{x})]$, is strictly increasing in its time allowance, S_{ij}.

LEMMA 1. *Constraints (8) can be replaced with the following:*

$$E[I_{ij}(\mathbf{S}_i, \mathbf{x})] = E[I_{i+1,j}(\mathbf{S}_{i+1}, \mathbf{x})], \quad 1 \le j \le i \le N-2. \tag{A.18}$$

Proof: Using a contradiction argument, one can show that $S_{ij} \ge \underline{d}_j = \inf d_j(\mathbf{x})$ (i.e., the lower support point of $d_j(\mathbf{x})$) in the optimal solution under both formulations (i.e., (8), as well as (A.18)). Therefore, we only need to show the equivalence with $S_{ij} \ge \underline{d}_j$ imposed, by induction. Note that, without loss of generality, we assume that for all $\epsilon > 0$, the probability of $d_j(\mathbf{x}) \in [\underline{d}_j, \underline{d}_j + \epsilon)$ is strictly positive. Otherwise, we can raise $\inf d_j(\mathbf{x})$ without affecting the distribution of $d_j(\mathbf{x})$. In addition, we temporarily drop $\mathbf{x}$ in $d_j(\mathbf{x})$, $I_{ij}(\mathbf{S}_i, \mathbf{x})$ and $B_{ij}(\mathbf{S}_i, \mathbf{x})$ to simplify the notation.

From the definitions of $I_{ij}(.)$ and $B_{ij}(.)$ in (9) and (10), one can observe that, for $k > j$, S_{ik} has no impact on $E[I_{ij}(\mathbf{S}_i)]$. Therefore, $E[I_{i1}(\mathbf{S}_i)]$ is only dependent on S_{i1} for all $i = 1, ..., N-1$. Because $E[I_{i1}(\mathbf{S}_i)]$ is strictly increasing in S_{i1} for $S_{i1} \ge \underline{d}_1$, and $E[I_{i-1,1}(\mathbf{S}_{i-1})]$ is strictly increasing in $S_{i-1,1}$ for $S_{i-1,1} \ge \underline{d}_1$, $S_{i1} = S_{i-1,1}$ is equivalent to $E[I_{i1}(\mathbf{S}_i)] = E[I_{i-1,1}(\mathbf{S}_{i-1})]$.

Suppose the statement "$S_{ik} = S_{i-1,k}$ for all pairs (i,k) where $k \le j$" is equivalent to "$E[I_{ik}(\mathbf{S}_i)] = E[I_{i-1,k}(\mathbf{S}_{i-1})]$ for all pairs (i,k) where $k \le j$". We want to show that the same holds for all pairs (i,k) where $k \le j+1$. Suppose $S_{ik} = S_{i-1,k}$ for all pairs (i,k) where $k \le j+1$. Since $S_{ik} = S_{i-1,k}$ for all pairs (i,k) with $k \le j$, (10) leads to that $B_{ij}(\mathbf{S}_i) = B_{i-1,j}(\mathbf{S}_{i-1})$ under any realization of $\mathbf{d}$. Therefore, $E[I_{i,j+1}(\mathbf{S}_i)] = E[(S_{i,j+1} - B_{ij}(\mathbf{S}_i) - d_{j+1})^+] = E[(S_{i-1,j+1} - B_{i-1,j}(\mathbf{S}_{i-1}) - d_{j+1})^+] = E[I_{i-1,j+1}(\mathbf{S}_{i-1})]$.

Next, we prove the reverse direction that $S_{i,j+1} > S_{i-1,j+1} \ge \underline{d}_{j+1}$ leads to $E[I_{i,j+1}(\mathbf{S}_i)] > E[I_{i-1,j+1}(\mathbf{S}_{i-1})]$. By induction assumption, $S_{ik} = S_{i-1,k}$ for all $k \le j$ indicates that $B_{ij}(\mathbf{S}_i) = B_{i-1,j}(\mathbf{S}_{i-1})$ under any realization of $\mathbf{d}$. Therefore, we have $(S_{i,j+1} - B_{ij}(\mathbf{S}_i) - d_{j+1})^+ \ge (S_{i-1,j+1} - B_{i-1,j}(\mathbf{S}_{i-1}) - d_{j+1})^+$. Next, we need to show that, there exists a positive probability $B_{ij}(\mathbf{S}_i) + d_{j+1} < S_{i,j+1}$. If that is the case, $P(S_{i,j+1} - B_{ij}(\mathbf{S}_i) - d_{j+1} > (S_{i-1,j+1} - B_{i-1,j}(\mathbf{S}_{i-1}) - d_{j+1})^+) > 0$. As a result, we will have $E[I_{i,j+1}(\mathbf{S}_i)] = E[(S_{i,j+1} - B_{ij}(\mathbf{S}_i) - d_{j+1})^+] > E[(S_{i-1,j+1} - B_{i-1,j}(\mathbf{S}_{i-1}) - d_{j+1})^+] = E[I_{i-1,j+1}(\mathbf{S}_{i-1})]$.

Now we show that such probability is strictly positive. For $k \le j$, let $\hat{E}_k$ denote the event that $d_k \le \underline{d}_k + \frac{S_{i,j+1} - d_{j+1}}{2N^2}$ and let $\hat{E} = \hat{E}_1 \wedge \hat{E}_2 \wedge ...\hat{E}_N$. Since the job durations are independent and $S_{i,j+1} > \underline{d}_{j+1}$, we know that $P(\hat{E}) > 0$. Define $\underline{\mathbf{d}} = (\underline{d}_1, ..., \underline{d}_{N-1})$. Then, conditioning on the event

$\hat{E}$, we have $P\left(B_{ij}(\mathbf{S}_i)+d_{j+1}\le B_{ij}(\underline{\mathbf{d}})+d_{j+1}|\hat{E}\right)=1$. Furthermore, the delay of the job in the j-th position is the largest for $\mathbf{d}\in\hat{E}$ if the realized duration of all jobs in all positions $k\le j$ achieve the respective upper bounds of $\underline{d}_k+\frac{S_{i,j+1}-\underline{d}_{j+1}}{2N^2}$. In such case, the worst-case delay of the job in the k-th position is $k\frac{S_{i,j+1}-\underline{d}_{j+1}}{2N^2}$, which implies that $P\left(B_{ij}(\mathbf{S}_i)+d_{j+1}\le(\frac{j^2+j}{2}+1)\frac{S_{i,j+1}-\underline{d}_{j+1}}{2N^2}+\underline{d}_{j+1}|\hat{E}\right)=1$. Note that $(\frac{j^2+j}{2}+1)\frac{S_{i,j+1}-\underline{d}_{j+1}}{2N^2}+\underline{d}_{j+1}<S_{i,j+1}$, as we know that $j\le N-1$ and we only need to consider $N\ge 2$, in which case $\frac{j^2+j}{2}+1<2N^2$. This implies $P\left(B_{ij}(\mathbf{S}_i)+d_{j+1}<S_{i,j+1}|\hat{E}\right)=1$, and hence, $P\left(B_{ij}(\mathbf{S}_i)+d_{j+1}<S_{i,j+1}\right)=P(\hat{E})>0$.

Thus, we have shown that $S_{i,j+1}>S_{i-1,j+1}\ge\underline{d}_{j+1}$ leads to $E[I_{i,j+1}(\mathbf{S}_i)]>E[I_{i-1,j+1}(\mathbf{S}_{i-1})]$. The proof of the equivalence between $S_{i,j+1}<S_{i-1,j+1}$ and $E[I_{i,j+1}(\mathbf{S}_i)]<E[I_{i-1,j+1}(\mathbf{S}_{i-1})]$ is also similar. □

After replacing (8) with (A.18), we perform Lagrangian relaxation on constraints (A.18), imposing Lagrangian multipliers λ_{ij}. For any $\{h_{ij}\}$ satisfying (6), and for any set of $\{\lambda_{ij}\}$ values, the Lagrangian function of (7) is given by:

$$\begin{aligned}L(\boldsymbol{\lambda}|\mathbf{h},\mathbf{x}) &= \min_{\mathbf{S}}\sum_{i=1}^{N-1}\left(\sum_{j=1}^{i}h_{ij}E[I_{ij}(\mathbf{S}_i,\mathbf{x})]+p_{i+1}(\mathbf{x})E[B_{ii}(\mathbf{S}_i,\mathbf{x})]\right)\\ &\quad+\sum_{i=1}^{N-2}\sum_{j=1}^{i}\lambda_{ij}\left(E[I_{ij}(\mathbf{S}_i,\mathbf{x})]-E[I_{i+1,j}(\mathbf{S}_{i+1},\mathbf{x})]\right)\\ &= \min_{\mathbf{S}}\sum_{i=1}^{N-1}\left(\sum_{j=1}^{i}(h_{ij}+\lambda_{ij}-\lambda_{i-1,j})E[I_{ij}(\mathbf{S}_i,\mathbf{x})]+p_i(\mathbf{x})E[B_{ii}(\mathbf{S}_i,\mathbf{x})]\right),\end{aligned}$$

where $\lambda_{j-1,j}=0$ and $\lambda_{N-1,j}=0$ for $j=1,...,N-1$. The Lagrangian dual is then given by $\max_{\boldsymbol{\lambda}}L(\boldsymbol{\lambda}|\mathbf{h},\mathbf{x})$, which gives a lower bound on the objective value of (7), given any $\mathbf{h}$ satisfying (6). Therefore, we may consider the problem

$$\max_{\mathbf{h}}\max_{\boldsymbol{\lambda}}L(\boldsymbol{\lambda}|\mathbf{h},\mathbf{x}) \quad \text{(A.19)}$$
$$\textit{subject to: } (6)$$

to obtain the best (largest) lower bound. In Proposition 2, we show that it is possible to combine $\mathbf{h}$ and $\boldsymbol{\lambda}$ into one single set of variables, and further decompose the resulting problem.

PROPOSITION 2. *The problems* (A.19) *and* (11) *are equivalent.*

Proof:

We first show that problem (A.19) is equivalent to $\max_{\mathbf{h}}\min_{\mathbf{S}}\sum_{i=1}^{N-1}\tilde{c}_i(\mathbf{S}_i,\mathbf{h},\mathbf{x})$, subject to (6), that is, the variables $\boldsymbol{\lambda}$ can be eliminated.

Given any set of $\boldsymbol{\lambda}$ values, we can uniquely compute a set of effective holding cost values, $\hat{h}_{ij}=(h_{ij}+\lambda_{ij}-\lambda_{i-1,j})$ where $\sum_{i=j}^{N-1}\hat{h}_{ij}=\kappa$ since $\lambda_{j-1,j}=0$ and $\lambda_{N-1,j}=0$ for $j=1,...,N-1$.

Conversely, given any $\hat{\mathbf{h}}$ and $\mathbf{h}$ where $\sum_{i=j}^{N-1}\hat{h}_{ij}=\sum_{i=j}^{N-1}h_{ij}=\kappa$, there exists a unique $\boldsymbol{\lambda}$ satisfying $\hat{h}_{ij}=(h_{ij}+\lambda_{ij}-\lambda_{i-1,j})$. To see that, define $\hat{\mathbf{h}}_j=(\hat{h}_{jj},...,\hat{h}_{N-2,j})^T$, $\mathbf{h}_j=(h_{jj},...,h_{N-2,j})^T$ and $\boldsymbol{\lambda}_j=(\lambda_{jj},...,\lambda_{N-2,j})^T$. Note that $\hat{h}_{N-1,j}$ and $h_{N-1,j}$ are determined uniquely by $\hat{\mathbf{h}}_j$ and $\mathbf{h}_j$ We observe that the balance equations, for any j, can be written in matrix form as $\hat{\mathbf{h}}_j-\mathbf{h}_j=A_j\boldsymbol{\lambda}_j$, where A_j is a $(N-1-j)\times(N-1-j)$ matrix. Let the component on the m-th row and the n-th column of matrix A_j be denoted by a_j^{mn}. Then:

$$a_j^{mn}=\begin{cases}1 & \text{if } m=n\\ -1 & \text{if } m=n+1\\ 0 & \text{otherwise.}\end{cases}$$

Therefore, the first row of A_j is given by $(1,0,...,0)$, the second row is given by $(-1,1,0,...,0)$, and the $(N-1-j)$-th row is given by $(0,0,...,-1,1)$. This matrix has full rank, and therefore the set of $\boldsymbol{\lambda}_j$ values satisfying $\hat{\mathbf{h}}_j-\mathbf{h}_j=A_j\boldsymbol{\lambda}_j$ is unique. In the remainder of this proof, we replace $\hat{\mathbf{h}}$ by $\mathbf{h}$ for notational brevity.

Next, we show that h_{ij} is nonnegative and nondecreasing in j in the optimal solution that maximizes $\tilde{C}(\mathbf{h},\mathbf{x})$.

First, we show that in the optimal $\mathbf{h}$, we will never have negative h_{ij} values. This is because, if $h_{ij}<0$ for any (i,j), the optimal solution in the inner minimization problem is to set $S_{ij}=\infty$, and the resulting value of $\min_{\mathbf{S}_i}\sum_{i=1}^{N-1}\tilde{c}_i(\mathbf{S}_i,\mathbf{h},\mathbf{x})=-\infty$. Therefore, such an $\mathbf{h}$ cannot be optimal in the maximization problem, and we may restrict h_{ij} to nonnegative values without affecting the optimal objective value.

Next, we claim that, along any chain i, there is an optimal solution in which h_{ij} is non-decreasing in j. To see that, suppose on some block i, we have $h_{i,j-1}>h_{ij}+\epsilon$ where $\epsilon>0$. Note that when $\mathbf{h}$ is fixed, (12) is the classical serial supply chain inventory problem (Clark and Scarf [5]). It is well known for this problem that, if a stage j has lower local holding cost than its upstream stage $j-1$, the resulting optimal local base stock level at stage $j-1$ must be 0, because there is no advantage of holding inventory at stage $j-1$ over stage j. This is true as long as $h_{i,j-1}\geq h_{ij}$, and within this range, the exact value of $h_{i,j-1}$ does not affect the expected costs. Recall that we are maximizing total expected costs of all chains over $\mathbf{h}$, subject to the constraint that $\sum_{i=1}^{N-1}h_{ij}=\kappa$. By reallocating ϵ to some other chain $k\neq i$ instead, we will not change the expected cost of chain i, while the cost of chain k will either increase or remain the same. Therefore, the total expected cost cannot decrease if one eliminates the possibility that $h_{i,j-1}>h_{ij}$. This implies that imposing the monotonicity constraint will not affect the optimal value of $\max_{\mathbf{h}}\min_{\mathbf{S}_i}\sum_{i=1}^{N-1}\tilde{c}_i(\mathbf{S}_i,\mathbf{h},\mathbf{x})$. In summary, we may simply restrict the feasible region of $\mathbf{h}$ as $\mathbf{H}$ without affecting the optimal objective value. $\square$

Online Appendix F: The Treatment of Overtime Cost and Session Length

In this Appendix, we consider the extension of including overtime costs and session length constraints in the model. Let T be the length of the session and p_{N+1} be the unit overtime cost. Let s_N be the time allowance for the job in the N-th position. Then $\sum_{i=1}^{N} s_j = T$, i.e., we require that the time allowances to be equal to the session length. In addition to (1) and (2), we add the following transition equations corresponding to the last job:

$$I_N(\mathbf{s},\mathbf{x}) = [s_N - d_N(\mathbf{x}) - B_{N-1}(\mathbf{s},\mathbf{x})]^+,$$
$$B_N(\mathbf{s},\mathbf{x}) = [s_N - d_N(\mathbf{x}) - B_{N-1}(\mathbf{s},\mathbf{x})]^-$$

By modifying (3) slightly, we can optimize the total expected idle and late-start penalty costs for a given sequence by solving the following problem.

$$\min_{\mathbf{s}\geq 0} \sum_{j=1}^{N} \kappa E\left[I_j(\mathbf{s},\mathbf{x})\right] + \sum_{j=1}^{N-1} p_{j+1}(\mathbf{x}) E\left[B_j(\mathbf{s},\mathbf{x})\right] + p_{N+1} B_N(\mathbf{s},\mathbf{x}) \tag{A.20}$$
$$\text{subject to} \quad \sum_{j=1}^{N} s_j = T. \tag{A.21}$$

To tackle this extension, we perform Lagrangian relaxation of constraint (A.21) by imposing a Lagrangian multiplier ζ. Then, the Lagrangian dual of (A.20) can be expressed as

$$\max_{\zeta} \min_{\mathbf{s}\geq 0} \sum_{j=1}^{N} \kappa E\left[I_j(\mathbf{s},\mathbf{x})\right] + \sum_{j=1}^{N-1} p_{j+1}(\mathbf{x}) E\left[B_j(\mathbf{s},\mathbf{x})\right] + p_{N+1} B_N(\mathbf{s},\mathbf{x}) + \zeta\left(\sum_{j=1}^{N} s_j - T\right). \tag{A.22}$$

Given any value of ζ, we now focus on the inner problem of (A.22). By Kong et al. [11], we have $\sum_{j=1}^{N} E\left[I_j(\mathbf{s},\mathbf{x})\right] = \sum_{j=1}^{N} s_j + E[B_N(\mathbf{s},\mathbf{x}) - \sum_{j=1}^{N} d_j(\mathbf{x})]$. Therefore, the inner problem of (A.22) is equivalent to

$$\min_{\mathbf{s}\geq 0} \sum_{j=1}^{N} (\kappa+\zeta) E\left[I_j(\mathbf{s},\mathbf{x})\right] + \sum_{j=1}^{N-1} p_{j+1}(\mathbf{x}) E\left[B_j(\mathbf{s},\mathbf{x})\right] + (p_{N+1}-\zeta) B_N(\mathbf{s},\mathbf{x}). \tag{A.23}$$

With the above treatment, one can see that Proposition 3 holds for the first $N-1$ positions for any $\zeta \in (-\kappa, p_{N+1})$, by going through the same steps of the proof. Note that optimal ζ cannot be less than $-\kappa$, because otherwise, the sum of total time allowances goes to infinity. Furthermore, if $\zeta > p_{N+1}$, it is *not* optimal to set $s_N > 0$, i.e., overtime should be scheduled deliberately. This can possibly happen when the session length is extremely short. Therefore, when the session length is sufficient such that the optimal $s_N > 0$, the main insight of this paper extends to the case with overtime cost and session length considerations. In addition, one can extend the OVP and OSP heuristics to this case by selecting among N sequences, generated by assigning each of the N job to the last position, and sequencing the remaining jobs in the first $N-1$ positions following OVP or OSP.

References

[1]W.P. Adams, R.J. Forrester and F.W. Glover, Comparisons and Enhancement Strategies for Linearizing Mixed 0-1 Quadratic Programs, *Discrete Optimization*, 1, 99-120, 2004.

[2]J.P. Aubin and I. Ekeland, *Applied Nonlinear Analysis*, John Wiley & Sons, New York, 1984.

[3]F. Chen, J.S. Song, Optimal policies for multi-echelon inventory problems with markov-modulated demand. *Operations Research* 49(2): 226-234, 2001.

[4]F. Chen, Y.S. Zheng, Lower bounds for multi-echelon stochastic inventory systems. *Management Science*, 40(11): 1426-1443, 994.

[5]A.J. Clark and H. Scarf, Optimal policies for a multi-echelon inventory problem. *Management science*, 6(4): 475–490, 1960.

[6]G. Gallego, J. Viapiano, and O. Ozer. A new algorithm and a new heuristic for serial supply systems. *Operations Research Letters*, 33(4), 349-362, 2005.

[7]D. Ge, G. Wan, Z. Wang and J. Zhang. A Convex Relaxation Approach to Appointment Scheduing. *Mathematics of Operations Research*, Forthcoming, 2013.

[8]P. Glasserman and S. Tayur. The stability of a capacitated, multi-echelon production-inventory system under a base-stock policy. *Operations Research*, 42(5), 913-925, 1994.

[9]W.T. Huh and G. Janakiraman, A sample-path approach to the optimality of echelon order-up-to policies in serial inventory systems. *Operations Research Letters*, 36(5): 547-550, 2008.

[10]W.T. Huh, G. Janakiraman and M. Nagarajan, capacitated serial inventory systems: sample path and stability properties under base-stock policies. *Operations Research*, 58(4), 1017-1022, 2010.

[11]Q. Kong, C. Lee, C.P. Teo and Z. Zheng. Scheduling Arrivals to a Stochastic Service Delivery System using Copositive Cones. *Operations Research*, 61(3): 711-726, 2013.

[12]K.H. Shang and J.S. Song, Newsvendor Bounds and Heuristic for Optimal Policies in Serial Supply Chains. *Management Science*, 49(5), 618-638, 2003.

[13]P. Zipkin. *Foundations of Inventory Management.* , McGraw-Hill, New York, 2000.