Appendix A. Proofs

Proof of Theorem 1. For all policies, the cost at DC j is

$$C_j^{\bullet}(S_0,S_j) = (h_j+c_j)\sum_{x=0}^{S_j-1}(S_j-x)P_j^{\bullet}(x)+c_jE(N_j^{\bullet})-c_jS_j,$$

establishing (4). We prove that $C_j^{\bullet}(S_0,S_j)$ is convex in S_j. Note that

$$C_j^{\bullet}(S_0,S_j)-C_j^{\bullet}(S_0,S_j-1)=(h_j+c_j)\sum_{x=0}^{S_j-1}P_j^{\bullet}(x)-c_j. \tag{A.29}$$

Since $\sum_{i=0}^{S_j-1}P_j^{\bullet}(i)$ is increasing in S_j and so does $C_j^{\bullet}(S_0,S_j)-C_j^{\bullet}(S_0,S_j-1)$. Therefore, for a given base-stock level at the warehouse, $C_j^{\bullet}(S_0,S_j)$ is convex in S_j and using (A.29) we get (5).

Proof of Lemma 1. Let $\tilde{w}_j^S(\cdot)$ denote the LT of the steady-state distribution of the total waiting time of an order in the system, i.e., the time elapsed between the instance that an order is placed by DC j and the instance that the corresponding product is received by the DC. We show that for the FCFS, SP, and MR policies

$$\Pi_j^{\bullet}(z)=\tilde{w}_j^S(\lambda_j(1-z)). \tag{A.30}$$

Note that the right hand side of (A.30) is the LT of the number of arrivals to DC j during the total waiting time of an order in the system, w_j^S, which includes the waiting time at the warehouse and transportation time. Therefore, we need to show that the distributional Little's law holds for our system under these three policies. We can apply Theorem 1 in Bertsimas and Nakazato (1995) directly. Let define the system as the manufacturer and transportation, i.e., an order enters the system as soon as a DC places an order and the order leaves the system when the DC receives its corresponding product. All required conditions given in Theorem 1 in Bertsimas and Nakazato (1995) hold for this system under these policies, i.e., 1) all arriving orders enter the system one at a time, remain in the system till served, and leave one at a time; 2) orders of each DC leave the system in order of their arrival (there is no order crossing based on our assumption); 3) new arriving orders of each DC do not affect the time in the system for previous orders of the DC. Therefore, we can apply Theorem 1 in Bertsimas and Nakazato (1995) directly to get (A.30). Since the waiting time of an order at the warehouse is independent of the transportation time of the corresponding product, we have

$$\tilde{w}_j^S(\cdot)=\tilde{w}_j(\cdot)\tilde{L}_j(\cdot), \tag{A.31}$$

where $\tilde{w}_j(\cdot)$ and $\tilde{L}_j(\cdot)$ are the LTs of the waiting time of an order at the warehouse and transportation time between the warehouse and DC j, respectively. Substituting (A.31) in (A.30), we get

$$\Pi_j^{\bullet}(z) = \tilde{w}_j^S(\lambda_j(1-z)) = \tilde{w}_j(\lambda_j(1-z))\tilde{L}_j(\lambda_j(1-z)) = \Pi_{L_j}^{\bullet}(z)\Pi_{W_j}^{\bullet}(z) \tag{A.32}$$

which completes the proof.

Proof of Theorem 2. The theorem follows from EC.1 in ABB (see also (B.39) in Appendix B): based on the construction of the FCFS backlog queue, the steady-state waiting time distribution of an order that finds the warehouse out of stock is identical in distribution to the steady-state waiting time distribution of a job in the FCFS backlog queue. This queue is an $M/G/1$ FCFS queue with an exceptional first service time in each busy period. Setting the low-priority arrival rate to zero in Theorem EC.1 in ABB, we get $\tilde{w}_j^{FC}(s)$ as in (12).

Proof of Lemma 2. To obtain $\Pi_j^{GMR}(z)$ under the GMR policy, we condition on the total number of orders in the warehouse, N_0:

$$\Pi_j^{GMR}(z) = \Pi_j^{GMR}(z|N_0 < -R_2+S_0)P(N_0 < -R_2+S_0) + \Pi_j^{GMR}(z|N_0 \geq -R_2+S_0)P(N_0 \geq -R_2+S_0). \tag{A.33}$$

To obtain $\Pi_j^{GMR}(z|N_0 < -R_2+S_0)$ we consider the third backlog queue, BQ_3. Note that this queue is a single class queue and therefore the number of people in it is independent of the allocation policy as long as the policy is non-idling. Therefore, to obtain the number of people in the system when the system is serving customers from BQ_3, we can consider a single class system working under the FCFS policy. Let us define the *third backlog system* as the third backlog queue plus the transportation time. In other words, the third backlog system is defined as the steady-state time elapsed between the instances that an order is placed by DC j given $N_0 < -R_2+S_0$ and the instance that the corresponding product is received by DC j when the manufacturer is in BQ_3. Then, similar to the proof of Lemma 1 we get

$$\Pi_j^{GMR}(z|N_0 < -R_2+S_0) = \Pi_{W_j^{BQ_3}}^{GMR}(z|N_0 < -R_2+S_0)\Pi_{L_j}(z), \tag{A.34}$$

where $\Pi_{W_j^{BQ_3}}^{GMR}(z|N_0 < -R_2+S_0)$ is the PGF of the outstanding number of orders in the third backlog queue given $N_0 < -R_2+S_0$.

To analyze $\Pi_j^{GMR}(z|N_0 \geq -R_2+S_0)$, we consider the second backlog queue, BQ_2, which is a two-priority queue and we get (similar to Lemma 1)

$$\Pi_j^{GMR}(z|N_0 \geq -R_2+S_0) = \Pi_{W_j^{BQ_2}}^{GMR}(z|N_0 \geq -R_2+S_0)\Pi_{L_j}(z). \tag{A.35}$$

Substituting (A.34) and (A.35) into (A.33), we get

$$\begin{aligned}
&\Pi_j^{GMR}(z)\\
&= \Pi^{GMR}_{W_j^{BQ_3}}(z|N_0 < -R_2+S_0)\Pi_{L_j}(z)P(N_0 < -R_2+S_0) + \Pi^{GMR}_{W_j^{BQ_2}}(z|N_0 \geq -R_2+S_0)\Pi_{L_j}(z)P(N_0 \geq -R_2+S_0)\\
&= \Pi_{L_j}(z)\left(\Pi^{GMR}_{W_j^{BQ_3}}(z|N_0 < -R_2+S_0)P(N_0 < -R_2+S_0) + \Pi^{GMR}_{W_j^{BQ_2}}(z|N_0 \geq -R_2+S_0)P(N_0 \geq -R_2+S_0)\right),
\end{aligned}$$

where

$$\Pi^{GMR}_{W_j}(z) = \Pi^{GMR}_{W_j^{BQ_3}}(z|N_0 < -R_2+S_0)P(N_0 < -R_2+S_0) + \Pi^{GMR}_{W_j^{BQ_2}}(z|N_0 \geq -R_2+S_0)P(N_0 \geq -R_2+S_0). \tag{A.36}$$

This completes the proof.

Proof of Theorem 3. Substituting (18), (19), and (22) in (20) we get (23).

Proof of Corollary 1. Note that the distribution of the total number of outstanding orders at DC j is the convolution of the distributions of number of the products on transit to DC j, and number of backlogs of DC j at the warehouse. Then, conditioning on whether an order arrival from DC j to the warehouse is served immediately or not, we get (24).

Appendix B. Relevant Analytical Results from ABB

Abouee-Mehrizi et al. (2012; from here on "ABB") consider a centralized multi-class $M/G/1$ make-to-stock queue and investigate the total cost of the system under the SP and MR policies. In this section we briefly explain how they analyze the system under these allocation polices and present some of their main results. We consider the warehouse as the centralized inventory system serving orders arriving from m classes of customers.

B.1. SP Policy in a Centralized Inventory System

ABB defined the SP Backlog (SPB) queue for the periods when the system is out of stock. They show that the SPB queue is a multi-priority $M/G/1$ queue with an exceptional first service time in a busy period as detailed below. Then, they demonstrate that the distribution of backlog in the SP system given it is out of stock is identical to the one in the SPB queue and characterize this distribution. Let S_0 denote the base-stock level of the system. Then, they consider the arrival process, service process, and the priority policy in the SPB queue as:

Arrival process: type r jobs arrive to the system according to a Poisson process with rate λ_r, $r = 1, 2, \ldots, m$.

Service process: the first service time in each busy period is exceptional and its distribution is identical to the equilibrium residual service time observed by arrivals who find exactly S_0 orders in the system upon their arrivals. We denote the LT of this exceptional service time by $\tilde{b}_{S_0}(\cdot)$. The distribution of the rest of the service times in each busy period is identical to the distribution of production times in the original system with a LT of $\tilde{b}(\cdot)$. Note that $\tilde{b}_{S_0}(\cdot)$ can be obtained recursively using, e.g., equation (4) in Kerner (2008) after setting $\lambda_i = \lambda$:

$$\tilde{b}_k(s) = \frac{\lambda}{s-\lambda}[\tilde{b}(\lambda)\frac{1-\tilde{b}_{k-1}(s)}{1-\tilde{b}_{k-1}(\lambda)} - \tilde{b}(s)], \quad k \geq 1, \tag{B.37}$$

where $\tilde{b}_0(s) = \tilde{b}(s)$ is the regular service time.

Priority policy: priority policy follows a static policy where type 1 has the highest priority and type m has the lowest priority, and the allocation decision is made at the service completion.

To derive the distribution of the number of type r jobs in the SPB queue, ABB observe that the utilization of this queue ,ρ_b, is:

$$\rho_b = \frac{\lambda\mu}{\mu_1\mu + \lambda(\mu - \mu_1)}, \tag{B.38}$$

where $1/\mu_1$ is the first moment of the exceptional first service times, $\tilde{b}_{S_0}(\cdot)$

Then, they characterize the LT of the steady-state waiting time distribution of type r job in the SPB queue and use it to derive the distribution of number of type r jobs in this queue. Let $\tilde{w}_j^{SP}(\cdot|I_0 = 0)$ denote the LT of the steady-state waiting time distribution of type r jobs in the SPB queue, $\lambda_j^+ = \sum_{i=1}^{j} \lambda_i$ and $\lambda_j^- = \sum_{i=j+1}^{m} \lambda_i$. Then,

$$\tilde{w}_j^{SP}(s|I_0 = 0) = \tilde{w}_h(s + \lambda_{j-1}^+(1 - \gamma_{j-1}^+(s))), \tag{B.39}$$

where

$$\begin{aligned}\tilde{w}_h(s) = & \frac{\tilde{b}(s)(1-\rho_b)(\lambda_l w_0 - \lambda) + (\pi_0 - (1-\rho_b))v_0\lambda(\tilde{b}(s) - 1)}{\lambda_h(1-\tilde{b}(s)) - s} \\ & + \frac{(1-\rho_b)(\tilde{b}_{S_0}(s)(\lambda - s) - \lambda_l w_0)}{\lambda_h(1-\tilde{b}(s)) - s}.\end{aligned} \tag{B.40}$$

and $\lambda_h = \lambda_j^+$, $\lambda_l = \lambda_j^-$, $v_0 = \tilde{b}(\lambda_h)$, $w_0 = \tilde{b}_{S_0}(\lambda_h)$, $\gamma_{j-1}^+(s) = \tilde{b}(s + \lambda_{j-1}^+(1 - \gamma_{j-1}^+(s)))$.

B.2. MR Policy in a Centralized Inventory System

To investigate the MR policy, ABB also considered a series of backlog queues to drive the total holding and backlog costs of the MR system. They define the r^{th} backlog queue, BQ_r, as a two-priority $M/G/1$ as detailed below.

Arrival process: the arrival process of the high and low-priority jobs to the system follows Poisson processes with rates of $\sum_{i=1}^{r-1} \lambda_i$ and λ_r, respectively.

Service process: the first service time in each busy period of the r^{th} backlog queue is exceptional and its distribution is identical to the equilibrium residual service time observed by high-priority arrivals to BQ_{r+1} who find exactly $\Delta_r := R_{r+1} - R_r$ high-priority jobs in the queue upon their arrivals. We denote the LT of this exceptional service time by $\tilde{b}^r_{\Delta_r}(s)(\cdot)$ and note that it can be obtained using Algorithm 1 in ABB. The distribution of the rest of the service times in each busy period is identical to production time in the original system with a LT of $\tilde{b}(\cdot)$.

Priority policy: at the service completion a low-priority job is served only if there are no high-priority jobs in the queue.

ABB obtain the distribution of number of type r backlogs in the system given that the inventory level is less than $R_r + 1$ using the LT of the steady-state waiting time distribution of type r in BQ_r:

$$\tilde{w}^{MR}_r(s|I_0 \leq R_r) = \tilde{w}^{BQ_r}_r(s) = \tilde{w}_h(s + \lambda_h(1 - \gamma^+_{r-1}(s))), \tag{B.41}$$

where

$$\tilde{w}_h(s) = \frac{\left(1-\rho^j_b\right)\left(\lambda_h\left(\tilde{b}(s) - \tilde{b}^r_{\Delta_r}(s)\right) + s\tilde{b}^r_{\Delta_r}(s)\right)}{s - \lambda_h\left(1 - \tilde{b}(s)\right)}, \tag{B.42}$$

and $\lambda_h = \lambda^+_{r-1}$, $\lambda_l = \lambda_r$.

ABB derive the distribution of the number high-priority jobs in BQ_r and obtain the distribution of the inventory level given that the inventory level is between R_r and R_{r-1} using the shortfall process of BQ_r. Let $P^{BQ_r}_h(i)$ denote the steady-state probability of having i high-priority jobs in BQ_r. Then,

$$P^{BQ_r}_h(i) = (1-\rho^r_b)\prod_{k=0}^{i-1}\frac{1-\tilde{b}^{r-1}_k(\lambda^+_{j-1})}{\tilde{b}(\lambda^+_{j-1})},\ i = 1, ... \tag{B.43}$$

where $\tilde{b}^{r-1}_k(\cdot)$ can be obtained using Algorithm 1 in ABB. Using (B.43), the probability that inventory level is less than R_{r+1} is:

$$\bar{F}^{MR}_r(R_r) = \prod_{j=r+1}^{m+1}\bar{F}^{BQ_j}_h(\Delta_j - 1) \tag{B.44}$$

where $\bar{F}^{BQ_j}_h(\Delta_j - 1) := 1 - \sum_{k=0}^{\Delta_j - 1} P^{BQ_k}_h(k)$.

References

Abouee-Mehrizi, H., B. Balcioglu, O. Baron 2012. “Strategies for a Single Product Multi-Class $M/G/1$ Queueing System”, *Operations Research*, Vol. 60, 803–812.

Kerner, Y. 2008. “The Conditional Distribution of the Residual Service Time in the $M_n/G/1$ Queue”, *Stochastic Models*, Vol. 24, 364–375.