Singapore Management University

Institutional Knowledge at Singapore Management University

Research Collection Lee Kong Chian School Of Business

Lee Kong Chian School of Business

6-2017

Closed-Form Approximations for Optimal (r, q) and (S, T) Policies in a Parallel Processing Environment

Marcus ANG
Singapore Management University, marcusang@smu.edu.sg

Karl SIGMAN
Columbia University

Jing-Sheng SONG
Duke University

Hanqin ZHANG
National University of Singapore
DOI: https://doi.org/10.1287/opre.2017.1623

Published in Operations Research, Vol 65 (5), Jun, 2017, pp. 1414-1428
https://doi.org/10.1287/opre.2017.1623
Author-final version

Closed-Form Approximations for Optimal (r, q) and (S, T) Policies in a Parallel Processing Environment

Marcus Ang • Karl Sigman • Jing-Sheng Song • Hanqin Zhang

Lee Kong Chian School of Business, Singapore Management University, 178899, Singapore
Department of IEOR, Columbia University, NY 10027, USA
The Fuqua School of Business, Duke University, Durham, NC 27708, USA
Business School, The National University of Singapore, 119245, Singapore

marcusang@smu.edu.sg • *ksigman@columbia.edu* • *jssong@duke.edu* • *bizzhq@nus.edu.sg*

July 28, 2013; revised December 30, 2014; October 22, 2015; April 6, 2016; August 2, 2016

Abstract

We consider a single-item continuous-review (r, q) inventory system with a renewal demand process and i.i.d. stochastic leadtimes. Using a stationary marked point process technique and a heavy traffic limit, we prove a previous conjecture that inventory position and inventory on-order are asymptotically independent. We also establish closed-form expressions for the optimal policy parameters and system cost in heavy traffic limit, the first of their kind to our knowledge. These expressions sharpen our understanding of the key determinants of the optimal policy and their quantitative and qualitative impacts. For example, the results demonstrate that the well-known square-root relationship between the optimal order quantity and demand rate under a sequential processing environment is replaced by the cube root under a stochastic parallel processing environment. We further extend the study to periodic-review (S, T) systems with constant leadtimes.

Keywords: inventory system, (r, q) policy, stochastic leadtime, asymptotic analysis, heavy-traffic limit.

1 Introduction

In this paper (with the exception of Section 6), we study a basic single-item continuous-review (r, q) inventory system, where r is the reorder point and q the order size. Both r and q are nonnegative integers. The demand follows a renewal process with rate λ. The replenishment leadtimes are independent, identically distributed (i.i.d.) random variables. Let L denote the generic random variable with the common distribution. All stockouts are backordered. There is a fixed order cost K for each order placed, and there are a linear inventory-holding cost with unit rate h and a linear backorder-penalty cost with unit rate p. The objective is to minimize the expected long-run average total system cost among all (r, q) policies. We denote the optimal policy by (r_*, q_*). (In general, we assume $K > 0$. When $K = 0$, we assume $q = 1$, so the policy reduces to a base-stock policy with base-stock level $r + 1$. For consistency, in this case, the optimal policy is denoted by $(r_*, 1)$, with $q_* = 1$.)

The (r, q) policy is widely used in practice and has received a lot of attention in the academic literature. (This form of policy is known to be suboptimal for systems with i.i.d. random leadtimes. The form of the optimal policy among all possible control policies is much more complex and remains unknown. See, for example, Zalkind 1978 and Benjaafar et al. 2014.) The early works in the literature focus on developing computationally efficient procedures for policy evaluation and optimization. While these procedures greatly advance the decision support systems for practice, they act as a "black box." That is, one can obtain the numerical values of key performance measures or optimal policy parameters after inputting the problem data, but these numbers cannot tell a "story", i.e., how the output depends on the input, such as the demand rate and leadtime variance. To overcome this shortcoming, more recent works strive to develop simple approximations to reveal the key determinants of system performance and optimal policy parameters. The focus of the current paper is in line with this latter effort. Below we briefly review what we know and don't know and then state our contributions in more detail.

1.1 Different Leadtime Models

The literature on (r, q) systems can be classified by how the replenishment leadtime and the corresponding supply subsystem are modeled. The supply subsystem can be an endogenous, exogenous sequential, or exogenous parallel processing system. Different leadtime models not only capture different characteristics of the real operating system but also affect the type of methodology applicable for analysis (see Zipkin 2000, Chapter 7).

In "endogenous" stochastic leadtime models, the orders generated from the inventory location under study comprise the primary workload of the supply subsystem. Consequently, the leadtime of a replenishment order is the sojourn time that order experienced in the supply subsystem, which depends on how many orders have already been sent to the supply system. For this reason, this type of inventory systems is often called *make-to-stock queues*.

In an "exogenous" leadtime model, the replenishment orders from the inventory location under study accounts only a negligible fraction of the workload of the supply subsystem and hence do not influence the dynamics of that subsystem. "Sequential" means the supply subsystem preserves the order sequence despite the stochastic variations of the leadtime.

The i.i.d. stochastic leadtimes assumed in the current paper is an exogenous model of leadtimes, because the leadtime experienced by a particular order does not depend on how many orders we have already placed. In contrast to the exogenous sequential model, however, this supply subsystem is a parallel processing system – it is equivalent to an infinite-server queueing system, in which the service time is precisely the leadtime. Here, orders can crossover, i.e., an order placed at an earlier time may arrive later than the current order. This model is suitable, for example, when the supply subsystem consists of alternative production sites (or suppliers) and/or alternative transportation routes, such as what one may expect when ordering online. The exogenous sequential and parallel supply systems intersect only when the leadtime is a constant.

1.2 Previous Results under Exogenous Sequential Leadtimes

Let t be the continuous time variable, $IN(t)$ the net inventory at time t, $IO(t)$ the outstanding orders, and $IP(t) = IN(t)+IO(t)$ the inventory position. Then, under an (r, q) policy, whenever $IP(t)$ reaches r, we immediately place an order of size q to bring $IP(t)$ back to $r + q$.

When leadtimes are exogenous and sequential, the following flow conservation law plays a critical role in analysis:

$$IN(t + L) = IP(t) - D(t, t + L], \quad t \geq 0, \tag{1}$$

where $D(t, t+L]$ is the cumulative demand in the interval $(t, t+L]$. (This expression is precise when L is a constant. When L is a random variable, we have a similar relationship in sample path. We use this form here for brevity.) Let IN, IP and D denote the steady-state limit of these random variables, we have

$$IN = IP - D. \tag{2}$$

It has been shown that IP is uniformly distributed in $\{r + 1, ..., r + q\}$ and IP and D are independent, see, e.g., Zipkin (1986) and Song (2000). Thus, to evaluate the performance of any given policy, one can simply employ (2). Federgruen and Zheng (1992) present an exact algorithm to find an optimal (r, q) policy. (More recently, Muthuraman et al. 2015 analyze a diffusion-process-type continuous demand model and obtain the optimality of the (s, S) policy and the limiting distribution of the inventory position for the discounted cost case. They also obtain the long run average system cost under any (s, S) policy.)

To better understand how system parameters affect the optimal policy, Zheng (1992) considers continuous approximations of the system, i.e., treating r and q as continuous variables. He relates q_* with the well-known EOQ formula and r_* with the newsvendor model. He shows that if the EOQ formula is used as a heuristic order quantity, the corresponding optimal reorder point can be computed as a newsvendor solution, and the resulting optimality loss is at most 12.5%. This error bound has subsequently been improved by Axsäter (1996) to $(\sqrt{5}-2)/2 \approx 11.8\%$ and by Gallego (1998) to 6.07% for a variant of the EOQ heuristic. Thus, it is generally understood that the optimal order quantity roughly grows in the square root of the mean demand rate and fixed order cost, i.e.,

$$q_* \sim O(\sqrt{K\lambda}), \tag{3}$$

as suggested by the EOQ formula. Ang et al. (2013) revisit these properties when r and q are restricted to integers.

Zheng (1992) shows that demand uncertainty drives q_* greater than the EOQ formula, but there is no quantification on how exactly demand variability affects q_*. Using stochastic comparison techniques, Song et al. (2010) investigate monotonicity properties of optimal policy parameters and system cost when leadtime or demand are stochastic larger or more variable. Federguren and Wang (2012) further study monotonic effect of general model primitives, including the cost parameters. These last two studies too do not quantify the effects.

One exception is Platt et al. (1997), who study a system with a *constant* leadtime L and assume the leadtime demand distribution is uniquely characterized by its mean λL and standard deviation σ, such as the normal distribution. These authors develop two closed-form heuristics for the optimal policy parameters under a constrained service level ι (fraction of demand satisfied from stock). One of the heuristic (the Simple Limit Case or SLC) has $q_* \sim O((1/\iota)\sqrt{K\lambda/h+\sigma^2})$, the other heuristic (the Atheoretic Heuristic or AH) has $q_* \sim O(\varpi(\iota)(K\lambda/h)^{1/3})$, where ϖ is a function related to the leadtime demand distribution. The latter is the only departure from the square-root relationship in the literature that we are aware of. The authors demonstrate numerically that AH performs better when the leadtime demand is normally distributed and the service level ι approaches to one.

1.3 Previous Results under Exogenous Parallel Leadtimes

Under the exogenous parallel processing environment (i.i.d. leadtimes), because orders can crossover, (2) no longer holds. From the definition of the inventory position $IP(t)$, however, we have

$$IN(t) = IP(t) - IO(t). \tag{4}$$

In steady state, we have

$$IN = IP - IO. \tag{5}$$

Thus, to evaluate a policy, we can employ (5). Because the policy dictates that every q demands generate an order, the supply subsystem is a $GI/GI/\infty$ queue. The difficulty here is, in general, IP and IO are not independent and their joint distribution relies on the interplay of the inventory and queuing subsystems.

Partly due to this difficulty, the literature on (r, q) system with i.i.d. leadtimes is relatively scant. For instance, we do not have an exact algorithm to find an optimal policy for the general system. Most of the existing works study performance measures for any given policy. For the special case of Poisson demand process and exponential L, Scarf (1958) and Galliher et al. (1959) obtain the exact distribution of IN in terms of transforms and intricate infinite series, respectively. Sahin (1983) extends these results to compound renewal demand and a more general L. More recently, Kulkarni and Yan (2012) analyze a system in which the demand rate changes according to a finite-state continuous-time Markov chain and the leadtimes are exponential. They use the matrix-geometric method to evaluate system performances.

To shed light on the determinants of system performance, Song and Zipkin (1996) develop two simple performance approximations for a system with Poisson demand and a general L, invoking (5). To use (5), they make two key assumptions: (i) IP is uniformly distributed and IO can be approximated by a normal distribution; (ii) IO and IP are independent. One of the normal distributions they employ is influenced by the heavy traffic limit in Whitt (1992). They conjecture that assumption (ii) is valid as λ grows large. They also "*expect that EOQ like effects govern the gross behavior of* q" so that "*the 'interesting' values of* q *are of order* $\sqrt{\lambda}$" (Song and Zipkin 1996, p.1356).

1.4 Our Contributions and Outline

In this paper, we extend Song and Zipkin (1996) in several important ways. First, we consider a general renewal demand process (see Section 2). Second, we prove that as λ goes to infinity, IP and IO are independent (Section 3). Third, we show that, IP converges in distribution to a uniform distribution and IO can be approximated by a normal distribution (Section 3). Thus, our results justify the key assumptions in Song and Zipkin (1996) even under a more general demand process.

More importantly, we examine the *optimal* policy and system behavior for this system. We obtain closed-form expressions for the optimal policy parameters and long-run average cost under a heavy-traffic limit (as λ gets larger); see Section 5. To the best of our knowledge, these expressions are the first of their kind for (r, q) inventory systems in general. Most strikingly, these results show that the well-known belief of (3) is true only if the leadtime is a *constant*; see Theorem 2 (B.i)-(B.iii). With general i.i.d. *random* leadtimes,

$$q_* = \left(\frac{2K}{C_*\sqrt{\nu}}\right)^{2/3} \lambda^{1/3} + o(\lambda^{1/3}), \tag{6}$$

where ν is a measure of leadtime variability and C_* is the optimal newsvendor cost with standard normal demand – a constant determined by the cost parameters p and h; see Theorem 2 (A.i)-(A.iii).

Furthermore, in Section 6, we develop similar asymptotic characteristics of a periodic-review (S, T) inventory system, where T is the review period and S the order-up-to level. Our efforts here join those by Bradley and Robinson (2005, 2008) in deriving closed-form approximations of inventory policy parameters for systems with i.i.d. leadtimes. These authors do not consider fixed order cost and focus on the periodic review, base-stock systems, where the review period is fixed and the base-stock level is optimized. In our study in Section 6, both review period and base-stock level are optimized. Thus, we study a more general system. In addition, while they establish bounds on the variance of outstanding orders, we employ an asymptotic analysis. Muharremoglu and Yang (2010) also consider periodic-review, base-stock systems without fixed order cost. They present a general exogenous leadtime model which includes the i.i.d. leadtimes and sequential leadtimes as special cases. Their focus, however, is on efficient method to compute the optimal base-stock level and cost, rather than on closed-form expressions.

Finally, the methods we use to derive these results may inspire similar approaches in the analysis of other inventory systems. Specifically, in Section 3, we first introduce the stationary marked point process technique to construct upper and lower bounds in the sense of stochastic orders, and use these bounds to establish the asymptotic independence of the inventory position and outstanding orders (Theorem 1). Then, with the help of the heavy traffic theory, we show that the outstanding orders after being properly centered and scaled converge to a normal distribution. In Section 4, we show the system cost of the original system converges to that of an auxiliary system with normally distributed demands. In Section 5, applying Taylor expansion to the first-order condition of the auxiliary model, we obtain the leading terms of the *optimal* policy parameters and system cost for the auxiliary system under high demand volume. We then argue that these leading terms are identical to those in the original system by showing the uniqueness of these terms in the auxiliary system. The analysis of the (S,T) system in Section 6 follows a similar procedure.

2 Notation and Preliminaries

We now introduce some additional notation and the detailed problem formulation. Let

$$
\begin{aligned}
t_n &= n\text{th demand arrival time},\\
\theta &= \text{coefficient of variation of the inter-demand time},\\
F(\cdot) &= \text{cumulative distribution function of } L,\\
1/\mu &= \text{mean leadtime} = \mathsf{E}[L],\\
\rho &= \lambda/\mu = \lambda\mathsf{E}[L] = \text{mean leadtime demand},\\
\eta &= \sqrt{\mu}/\theta,\\
L_{(2)} &= \min\{L^1, L^2\},\ \text{where } L^1 \text{ and } L^2 \text{ are two independent copies of } L,\\
\nu &= \mathsf{E}[L] - \mathsf{E}[L_{(2)}] = \frac{1}{2}\mathsf{E}[|L^1 - L^2|] = \frac{1}{\mu} - \int_0^\infty (1 - F(t))^2 \mathsf{d}t.
\end{aligned}
$$

Note that ν (≥ 0) is a measure of leadtime variability; it equals zero when the leadtime is deterministic.

To ease analysis, define

$$
\begin{aligned}
J(t) &= r + q - IP(t),\\
N(t) &= \text{number of outstanding orders in the supply system} = IO(t)/q,\\
\mathcal{Q} &= \{0, 1, 2, \ldots, q-1\} = \text{range of } J(t).
\end{aligned}
$$

Let N and J denote the random variables having the corresponding limiting distributions of $N(t)$ and $J(t)$. Then, by (5), in time-stationary,

$$IO = qN, \quad IN = r + q - J - qN. \tag{7}$$

For simplicity and without loss of generality, we assume the system starts with $IP(0) = r+q$. Then, $J(t) \in \mathcal{Q}$ acts as a counter process: It starts with zero and increases by 1 at each demand until it reaches q, at which moment we immediately place an order of size q and reset it back to zero.

Because we place the nth order at t_{nq}, $A_q(t) = \max\{n : t_{nq} \le t\}$ is the total number of orders placed by time t. Thus, the supply process can be viewed as a $GI/GI/\infty$ queue with arrival process $\{A_q(t), t \ge 0\}$ and service time distribution $F(\cdot)$. The arrivals process $\{A_q(t), t \ge 0\}$ is a q-phase renewal process in which the interarrival time has q phases, each one with a rate λ. The process $J(t)$ traces its phases precisely.

From the elementary renewal theory (see Theorem 3.3.4 on p.107, Ross 1996), we have $\lim_{T\to\infty} \mathsf{E} A_q(T)/T = \lambda/q$. Define

$$\hat{G}(y) = h \cdot (y)^+ + p \cdot (y)^-, \tag{8}$$

where y is any real number, $(y)^+ = \max\{0, y\}$, $(y)^- = \max\{0, -y\}$. Our objective is to minimize the expected long-run average system cost

$$\mathcal{AC}(r,q) = \lim_{T\to\infty} \frac{1}{T} \mathsf{E}\left(K \cdot A_q(T) + \int_0^T \hat{G}(IN(t))\mathsf{d}t\right) = \frac{\lambda K}{q} + \mathsf{E}[\hat{G}(IN)]. \tag{9}$$

Here, we assume $K > 0$. When $K = 0$, as mentioned above, we assume $q = 1$. In this case, the expected long-run average system cost is $\mathcal{AC}(r, 1) = \mathsf{E}[\hat{G}(IN)]$. For a proof of (9), see the Appendix.

Clearly, to solve the optimization problem (9), it is important to know the distribution on IN, which, in turn, is determined by the joint distribution of IP and IO. In general, this joint distribution is difficult to obtain. This is because $IP(t)$ and $IO(t)$ are correlated for any given t, so IP and IO may also be dependent. For tractability, we seek to study the asymptotic behavior of the system as λ approaches to infinity.

It turns out the asymptotic analysis critically depends on whether the replenishment lead-time L is a random variable ($\nu > 0$) or a constant ($\nu = 0$). For the special case when L is a constant $1/\mu$, as mentioned in the introduction, we can use an alternative relation (2) to obtain the distribution of IN. The corresponding cost function can be written as

$$\mathcal{AC}(r,q) = \frac{1}{q}\Big(\lambda K + \sum_{\ell=r+1}^{r+q} G(\ell)\Big), \tag{10}$$

$$G(\ell) = \mathsf{E}[\hat{G}(\ell - D)], \tag{11}$$

where D is the time-stationary of $D(t, t+1/\mu]$, the sum of demands that occur during the time interval $(t,\ t+1/\mu]$.

3 Asymptotic Behavior of IP and IO

Consider a sequence of the inventory systems with *random leadtimes* (i.e., $\nu > 0$) indexed by the demand rate λ, denoted as System-$\mathcal{S}^\lambda$. Consequently, all quantities introduced above will be superscripted by λ; e.g., the optimal policy is denoted by $(r_*^\lambda, q_*^\lambda)$. In this section, we will focus on the asymptotic behavior $(IP^\lambda(t), IO^\lambda(t))$, or, equivalently, that of $(J^\lambda(t), N^\lambda(t))$. To do so, we first let $t \to \infty$ and look at the steady-state limit for each component, i.e., the marginal distributions of J^λ and N^λ. We then study the asymptotic properties of joint distribution of (J^λ, N^λ) as $\lambda \to \infty$.

Let $\{\xi_k : k \geq 1\}$ and $\{\zeta_k : k \geq 1\}$ be independent i.i.d. sequences of nonnegative random variables with

$$\mathsf{E}\xi_1 = 1, \quad \text{and} \quad \mathsf{E}(\xi_1 - 1)^2 = \theta^2 < \infty; \tag{12}$$

$$\mathsf{E}\zeta_1 = \mu^{-1} < \infty, \quad \text{and} \quad \mathsf{E}(\zeta_1 - \mu^{-1})^2 < \infty. \tag{13}$$

For the λth system, define

$$t_0^\lambda = 0,\ t_k^\lambda - t_{k-1}^\lambda = \frac{\xi_k}{\lambda},\ k = 1, 2, \cdots,$$

and the kth order's leadtime is given by ζ_k with the distribution $F(\cdot)$.

Before analyzing the joint steady-state behavior of $J^\lambda(t)$ and $IO^\lambda(t)$, we look at their marginal steady-state behavior. First, the distribution of the steady-state limit of $J^\lambda(t)$ (and hence $IP^\lambda(t)$) directly follows from Simon (1968) (see p.6, Theorem).

Lemma 1 *For any $j \in \mathcal{Q}^\lambda$,*

$$\mathsf{Pr}\left(J^\lambda \leq j\right) = \lim_{t\to\infty} \mathsf{Pr}\left(J^\lambda(t) \leq j\right) = \frac{j+1}{q^\lambda}.$$

Next, we show that N^λ, and hence IO^λ, is approximately normally distributed as λ grows large. We do so by showing in the following lemma that an appropriately normalized and centered N^λ has an asymptotic standard normal distribution. To get the normalized and centered factors of N^λ, define (for random leadtimes)

$$\left(\sigma^\lambda(q^\lambda)\right)^2 = \frac{1}{\mu} + \Big(\frac{\theta^2}{q^\lambda} - 1\Big)\int_0^\infty (1 - F(t))^2 \mathsf{d}t = \nu + \frac{\theta^2}{q^\lambda}\mathsf{E}[L_{(2)}], \tag{14}$$

$$\gamma^\lambda(q^\lambda) = \sigma^\lambda\sqrt{\lambda q^\lambda}, \qquad \beta^\lambda(q^\lambda) = \frac{q^\lambda}{\gamma^\lambda(q^\lambda)}, \tag{15}$$

$$z^\lambda(i, q^\lambda) = \frac{i - \rho^\lambda}{\gamma^\lambda(q^\lambda)}, \qquad Y^\lambda(q^\lambda) = \frac{q^\lambda N^\lambda - \rho^\lambda}{\gamma^\lambda(q^\lambda)} = \beta^\lambda(q^\lambda)\Big(N^\lambda - \frac{\rho^\lambda}{q^\lambda}\Big). \tag{16}$$

By noting that ρ^λ is the mean leadtime demand, $(i-\rho^\lambda)$ just represents the net inventory after the leadtime if the inventory position is i. Thus $z^\lambda(i,q^\lambda)$ is a scaled net inventory level. Similarly, $(q^\lambda N^\lambda-\rho^\lambda)$ measures the fluctuation of the outstanding orders around the mean leadtime demand, and $Y^\lambda(q^\lambda)$ is a scaled and centered outstanding orders. The following condition will be useful:

Condition 1 $\lim_{\lambda\to\infty} q^\lambda/\lambda = 0$ *for the sequence of order sizes* $\{q^\lambda\}$.

Condition 1 gives a comparability relationship between the order size q^λ and demand rate λ. Let $\Phi(\cdot)$ be the standard normal distribution function. According to Theorem 1 in Borovkov (1967) or Proposition 2.5 in Whitt (1992), we have

Lemma 2 *If the sequence of order sizes* $\{q^\lambda\}$ *for System-*$\mathcal{S}^\lambda$ *satisfies Condition* 1*, then for any real number* y*,*

$$\lim_{\lambda\to\infty} \mathsf{Pr}\left(Y^\lambda(q^\lambda)\le y\right)=\Phi(y).$$

With the above lemmas, we now proceed to establish the following theorem on the asymptotic independence between the outstanding orders and inventory position. This result also justifies the key assumptions in Song and Zipkin (1996).

Theorem 1 *Consider System-*$\mathcal{S}^\lambda$*. For* $y\in(-\infty,+\infty)$ *and* $x^\lambda\in\mathcal{Q}^\lambda$*,*

$$\frac{1}{q^\lambda}\mathsf{Pr}\left(Y^\lambda(q^\lambda)\le y-\beta^\lambda(q^\lambda)\right)\le \mathsf{Pr}\left(Y^\lambda(q^\lambda)\le y, J^\lambda=x^\lambda\right)\le \frac{1}{q^\lambda}\mathsf{Pr}\left(Y^\lambda(q^\lambda)\le y+\beta^\lambda(q^\lambda)\right).$$

Moreover, if Condition 1 *holds, then* J^λ *and* N^λ *are asymptotically independent. That is,*

$$\lim_{\lambda\to\infty}\left[\mathsf{Pr}\left(Y^\lambda(q^\lambda)\le y, J^\lambda\le x^\lambda\right)-\Phi(y)\times\frac{x^\lambda+1}{q^\lambda}\right]=0.$$

Proof : First, for each λ, we examine the joint distribution of N^λ and J^λ and derive its upper and lower bounds. The difficulty here is that N^λ is already the steady-state of the process $\{N^\lambda(t):t\ge 0\}$ and J^λ already the steady-state of the process $\{J^\lambda(t):t\ge 0\}$. To obtain the joint distribution, we adopt the time-stationary point process framework discussed in Sigman (1996) to establish a sample-path relationship of the two random variables by constructing the two-sided versions of the original processes. Specifically, we consider the two-side infinite sequence $\{\tilde{t}_k^\lambda : k = \pm1, \pm2, \cdots\}$ with the following properties:

$$\begin{array}{ll} i) & \tilde{t}_{-1}^\lambda \le 0 < \tilde{t}_1^\lambda; \\ ii) & \{(\tilde{t}_k^\lambda - \tilde{t}_{k-1}^\lambda) : k = \pm1, \pm2, \cdots\} \text{ is a doulbe i.i.d. sequence;} \\ iii) & \mathsf{Pr}(-\tilde{t}_{-1}^\lambda \le x) = \mathsf{Pr}(\tilde{t}_1^\lambda \le x) = \dfrac{1}{\mathsf{E}(\tilde{t}_2^\lambda - \tilde{t}_1^\lambda)} \displaystyle\int_0^x \Big(1 - \mathsf{Pr}(\tilde{t}_2^\lambda - \tilde{t}_1^\lambda \le y)\Big)\,\mathsf{d}y. \end{array}$$

The marked points from $\{\tilde{t}_k^\lambda : k = \pm1, \pm2, \cdots\}$ are given by the following:

$\{\tilde{t}_{-(i+1+nq)}^\lambda : n \ge 0\}$ and $\{\tilde{t}_{-i+(n+1)q}^\lambda : n \ge 0\}$ are marked with probability $\dfrac{1}{q}$, $i = 0, 1, \cdots, q-1$.

Thus, analogous to the four-tuple process $(\psi_M^*, \psi_q^*, J^*, \psi^*)$ as described in Section 6.3 of Sigman and Whitt (2011), we have generated the four-tuple, two-side, jointly time-stationary process $(\psi_M^\lambda, \psi_q^\lambda, J^\lambda, \psi^\lambda)$ with

$$J^\lambda(t) = \begin{cases} q \| (J^\lambda(0) + R^\lambda(t)) \text{ with } R^\lambda(t) = \min\{k \le -1 : \tilde{t}_k^\lambda \ge t\}, & \text{if } t \le 0, \\ q \| (J^\lambda(0) + R^\lambda(t)) \text{ with } R^\lambda(t) = \max\{k \ge 1 : \tilde{t}_k^\lambda \le t\}, & \text{if } t > 0, \end{cases}$$

where "$\|$" is the modulo operator. When $\{\tilde{t}_k^\lambda : k = \pm1, \pm2, \cdots\}$ are considered as the demand arrival points, $(\tilde{t}_2^\lambda - \tilde{t}_1^\lambda)$ and $(t_2^\lambda - t_1^\lambda)$ follow the same distribution, the marked points $\{\tilde{t}_{-(i+1+nq)}^\lambda : n \ge 0\}$ and $\{\tilde{t}_{q-i+nq}^\lambda : n \ge 0\}$ trigger orders, and the nth order leadtime is experienced by $\{\zeta_n : n = \pm1, \pm2, \cdots\}$. From the time-stationary point process framework, we have

$$(J^\lambda(0), N^\lambda(0)) \text{ and } (J^\lambda, N^\lambda) \text{ have the same distribution.} \tag{17}$$

To prove the first part of the theorem, it is sufficient to consider the joint distribution of $(J^\lambda(0), N^\lambda(0))$. First, the relationship between $J^\lambda(0)$ and $N^\lambda(0)$ can be easily established by observing the following fact:

$$\text{if } J^\lambda(0) = j \in \mathcal{Q}^\lambda, \text{ then } N^\lambda(0) = \sum_{n=1}^{\infty} I\Big\{-\tilde{t}_{-(1+j+(n-1)q^\lambda)}^\lambda < \zeta_{-n}\Big\}, \tag{18}$$

where $I\{A\}$ is the indicator function of event A. From (18), we observe that $N^\lambda(0)$ depends on $J^\lambda(0)$. However, we next show that there exist upper and lower bounds on $N^\lambda(0)$ which are independent of $J^\lambda(0)$. Moreover, the difference between the upper and lower bounds is bounded by a constant (independent of λ). Thus, the dependence of $N^\lambda(0)$ on $J^\lambda(0)$ will gradually disappear as λ grows large.

Consider any sample path ω. To simplify notation, we suppress the notation ω in the following sample-path argument. In other words, the statement about random variables hold with probability one. Noting that for $j \in \mathcal{Q}$, $\tilde{t}^{\lambda}_{-(1+j+(n-1)q^{\lambda})}$ and $\tilde{t}^{\lambda}_{-(1+(n-1)q^{\lambda})}$ are the $(1 + j + (n-1)q^{\lambda})$th and $(1 + (n-1)q^{\lambda})$th demand arrivals counting back from time zero, and $1 + j + (n-1)q^{\lambda} \geq 1 + (n-1)q^{\lambda}$, we have

$$\tilde{t}^{\lambda}_{-(1+j+(n-1)q^{\lambda})} \leq \tilde{t}^{\lambda}_{-(1+(n-1)q^{\lambda})}, \quad j \in \mathcal{Q}^{\lambda} \text{ and } n = 1, 2, \cdots. \tag{19}$$

Therefore, for $j \in \mathcal{Q}^{\lambda}$,

$$\sum_{n=1}^{\infty} I\Big\{-\tilde{t}^{\lambda}_{-(1+j+(n-1)q^{\lambda})} < \zeta_{-n}\Big\} \leq \sum_{n=1}^{\infty} I\Big\{-\tilde{t}^{\lambda}_{-(1+(n-1)q^{\lambda})} < \zeta_{-n}\Big\}. \tag{20}$$

Furthermore, using (19),

$$\sum_{n=1}^{\infty} I\Big\{-\tilde{t}^{\lambda}_{-(1+j+nq^{\lambda})} < \zeta_{-n}\Big\} \leq \sum_{n=1}^{\infty} I\Big\{-\tilde{t}^{\lambda}_{-(1+nq^{\lambda})} < \zeta_{-n}\Big\}. \tag{21}$$

Noting that $I\Big\{-\tilde{t}^{\lambda}_{-(1+j)} < \zeta_1\Big\} \leq 1$, by (21), we obtain

$$I\Big\{-\tilde{t}^{\lambda}_{-(1+j)} < \zeta_1\Big\} + \sum_{n=1}^{\infty} I\Big\{-\tilde{t}^{\lambda}_{-(1+j+nq^{\lambda})} < \zeta_{-n}\Big\} \leq 1 + \sum_{n=1}^{\infty} I\Big\{-\tilde{t}^{\lambda}_{-(1+nq^{\lambda})} < \zeta_{-n}\Big\}. \tag{22}$$

By the fact that $\{\zeta_n : n = \pm 1, \pm 2, \cdots\}$ is i.i.d. and is independent of $\{\tilde{t}^{\lambda}_k : k = \pm 1, \pm 2, \cdots\}$ (as the demand arrivals and leadtimes are independent), we know that

$$\sum_{n=1}^{\infty} I\Big\{-\tilde{t}^{\lambda}_{-(1+j+(n-1)q^{\lambda})} < \zeta_{-n}\Big\} \text{ and } I\Big\{-\tilde{t}^{\lambda}_{-(1+j)} < \zeta_1\Big\} + \sum_{n=1}^{\infty} I\Big\{-\tilde{t}^{\lambda}_{-(1+j+nq^{\lambda})} < \zeta_{-n}\Big\}$$

have the same distribution for $j \in \mathcal{Q}^{\lambda}$. Thus, it follows from (22) that for $j \in \mathcal{Q}^{\lambda}$,

$$\sum_{n=1}^{\infty} I\Big\{-\tilde{t}^{\lambda}_{-(1+j+(n-1)q^{\lambda})} < \zeta_{-n}\Big\} \leq_{\text{s.t.}} 1 + \sum_{n=1}^{\infty} I\Big\{-\tilde{t}^{\lambda}_{-(1+nq^{\lambda})} < \zeta_{-n}\Big\}. \tag{23}$$

Applying (18) and (23) yields

$$N^{\lambda}(0) \leq_{\text{s.t.}} 1 + \sum_{n=1}^{\infty} I\Big\{-\tilde{t}^{\lambda}_{-(1+nq^{\lambda})} < \zeta_{-n}\Big\}. \tag{24}$$

Thus, we obtain an upper bound on $N^{\lambda}(0)$ in the sense of stochastic orders, which is independent of $J^{\lambda}(0)$.

Symmetrically, similar to (19), we have

$$\tilde{t}^{\lambda}_{-(1+j+(n-1)q^{\lambda})} \geq \tilde{t}^{\lambda}_{-(1+nq^{\lambda})}, \; j \in \mathcal{Q}^{\lambda} \text{ and } n = 1, \cdots. \tag{25}$$

This implies $I\Big\{-\tilde{t}^{\lambda}_{-(1+j+(n-1)q^{\lambda})} < \zeta_{-n}\Big\} \geq I\Big\{-\tilde{t}^{\lambda}_{-(1+nq^{\lambda})} < \zeta_{-n}\Big\}$. Hence, for $j \in \mathcal{Q}^{\lambda}$,

$$\sum_{n=1}^{\infty} I\Big\{-\tilde{t}^{\lambda}_{-(1+j+(n-1)q^{\lambda})} < \zeta_{-n}\Big\} \geq \sum_{n=1}^{\infty} I\Big\{-\tilde{t}^{\lambda}_{-(1+nq^{\lambda})} < \zeta_{-n}\Big\}. \tag{26}$$

Along the same line of the proof of (24), by (25)-(26), we can prove

$$\sum_{n=1}^{\infty} I\Big\{-\tilde{t}^{\lambda}_{-(1+(n-1)q^{\lambda})} < \zeta_{-n}\Big\} \leq_{\mathsf{s.t.}} 1 + N^{\lambda}(0). \tag{27}$$

Thus, we also obtain a lower bound on $N^{\lambda}(0)$ (in the sense of stochastic orders) that is independent of $J^{\lambda}(0)$.

Similar to $Y^{\lambda}(q^{\lambda})$ given by (16), let $Y_0^{\lambda}(q^{\lambda}) = \beta^{\lambda}(q^{\lambda}) \cdot (N^{\lambda}(0) - \rho^{\lambda}/q^{\lambda})$. We have

$$\begin{aligned}
&\mathsf{Pr}\Big(J^{\lambda}(0) = x^{\lambda},\ \ Y_0^{\lambda}(q^{\lambda}) \leq y\Big) \\
&= \mathsf{Pr}\Big(J^{\lambda}(0) = x^{\lambda}, \beta^{\lambda}(q^{\lambda}) \cdot \Big(\sum_{n=1}^{\infty} I\Big\{-\tilde{t}^{\lambda}_{-(1+x^{\lambda}+(n-1)q^{\lambda})} < \zeta_{-n}\Big\} - \frac{\rho^{\lambda}}{q^{\lambda}}\Big) \leq y\Big) \quad \text{(by (18))} \\
&\leq \mathsf{Pr}\Big(J^{\lambda}(0) = x^{\lambda}, \beta^{\lambda}(q^{\lambda}) \cdot \Big(\sum_{n=1}^{\infty} I\Big\{-\tilde{t}^{\lambda}_{-(1+nq^{\lambda})} < \zeta_{-n}\Big\} - \frac{\rho^{\lambda}}{q^{\lambda}}\Big) \leq y\Big) \quad \text{(by (26))} \\
&= \mathsf{Pr}\Big(J^{\lambda}(0) = x^{\lambda}\Big) \times \mathsf{Pr}\Big(\beta^{\lambda}(q^{\lambda}) \cdot \Big(\sum_{n=1}^{\infty} I\Big\{-\tilde{t}^{\lambda}_{-(1+nq^{\lambda})} < \zeta_{-n}\Big\} - \frac{\rho^{\lambda}}{q^{\lambda}}\Big) \leq y\Big) \\
&= \frac{1}{q^{\lambda}} \times \mathsf{Pr}\Big(\beta^{\lambda}(q^{\lambda}) \cdot \Big(\sum_{n=1}^{\infty} I\Big\{-\tilde{t}^{\lambda}_{-(1+nq^{\lambda})} < \zeta_{-n}\Big\} - \frac{\rho^{\lambda}}{q^{\lambda}}\Big) \leq y\Big) \quad \text{(by Lemma 1)} \\
&\leq \frac{1}{q^{\lambda}} \times \mathsf{Pr}\Big(Y_0^{\lambda}(q^{\lambda}) - \beta^{\lambda}(q^{\lambda}) \leq y\Big) \quad \text{(by (24))} \\
&= \frac{1}{q^{\lambda}} \times \mathsf{Pr}\Big(Y_0^{\lambda}(q^{\lambda}) \leq y + \beta^{\lambda}(q^{\lambda})\Big),
\end{aligned} \tag{28}$$

where the second equality follows from the observations below:

a) $J^{\lambda}(0)$ is independent of $\{\tilde{t}^{\lambda}_k : k \leq -1\}$ (Theorem 8 in Sigman and Whitt 2011);

b) $J^{\lambda}(0)$ is independent of the sequence $\{\zeta^{(1)}_k : k \leq -1\}$ defined by

$$\zeta^{(1)}_k = \begin{cases} 0, & \text{if } k \neq 1 + nq, \\ \zeta_n, & \text{if } n = 1 + nq \end{cases} \quad \text{(as the leadtimes are independent of the demand arrivals);}$$

c) $\sum_{k=1}^{\infty} I\Big\{-\tilde{t}^{\lambda}_{-k} < \zeta^{(1)}_{-k}\Big\} = \sum_{n=1}^{\infty} I\Big\{-\tilde{t}^{\lambda}_{-(1+nq^{\lambda})} < \zeta_{-n}\Big\}$.

Analogously,

$$\mathsf{Pr}\Big(J^{\lambda}(0) = x^{\lambda},\ \ Y_0^{\lambda}(q^{\lambda}) \leq y\Big)$$

$$\geq \frac{1}{q^\lambda} \times \mathsf{Pr}\Big(\beta^\lambda(q^\lambda) \cdot \Big(\sum_{n=1}^{\infty} I\Big\{ -\tilde{t}^\lambda_{-(1+(n-1)q^\lambda)} < \zeta_{-n}\Big\} - \frac{\rho^\lambda}{q^\lambda}\Big) \leq y\Big) \quad \text{(by (20))}$$

$$\geq \frac{1}{q^\lambda} \times \mathsf{Pr}\Big(Y_0^\lambda(q^\lambda) + \beta^\lambda(q^\lambda) \leq y\Big) \quad \text{(by (27))}$$

$$= \frac{1}{q^\lambda} \times \mathsf{Pr}\Big(Y_0^\lambda(q^\lambda) \leq y - \beta^\lambda(q^\lambda)\Big). \tag{29}$$

Combining (28)-(29), we obtain the first part of the theorem.

Note that from the first part of the theorem we have

$$\begin{aligned} \frac{x^\lambda + 1}{q^\lambda}\mathsf{Pr}\left(Y^\lambda(q^\lambda) \leq y - \beta^\lambda(q^\lambda)\right) &\leq \mathsf{Pr}\left(Y^\lambda(q^\lambda) \leq y, J^\lambda \leq x^\lambda\right) \\ &\leq \frac{x^\lambda + 1}{q^\lambda}\mathsf{Pr}\left(Y^\lambda(q^\lambda) \leq y + \beta^\lambda(q^\lambda)\right). \end{aligned} \tag{30}$$

Therefore,

$$\begin{aligned} &\mathsf{Pr}\left(Y^\lambda(q^\lambda) \leq y, J^\lambda \leq x^\lambda\right) - \frac{x^\lambda + 1}{q^\lambda}\Phi(y) \\ &\leq \frac{x^\lambda + 1}{q^\lambda} \times \max\Big\{\Big|\mathsf{Pr}\left(Y^\lambda(q^\lambda) \leq y + \beta^\lambda(q^\lambda)\right) - \Phi(y)\Big|, \Big|\mathsf{Pr}\left(Y^\lambda(q^\lambda) \leq y - \beta^\lambda(q^\lambda)\right) - \Phi(y)\Big|\Big\}. \end{aligned} \tag{31}$$

Then the second part follows directly from

$$\begin{aligned} \frac{x^\lambda + 1}{q^\lambda} &\leq 1 \quad (\text{by } x^\lambda \in \mathcal{Q}^\lambda), \\ \lim_{\lambda\to\infty} \mathsf{Pr}\Big(Y^\lambda(q^\lambda) \leq y + \beta^\lambda(q^\lambda)\Big) &= \lim_{\lambda\to\infty} \mathsf{Pr}\Big(Y^\lambda(q^\lambda) \leq y - \beta^\lambda(q^\lambda)\Big) = \Phi(y), \end{aligned}$$

where Lemma 2 and the fact that $\lim_{\lambda\to\infty} \beta^\lambda(q^\lambda) = 0$ (implied by the assumption $\lim_{\lambda\to\infty} q^\lambda/\lambda = 0$) are used. ■

4 Auxiliary Systems

Given that under random leadtimes, $Y^\lambda(q^\lambda)$ (and thus N^λ) approaches to a normally distributed random variable as λ goes to infinity, in this section we show that the long-run average system cost (of the original system) converges to its continuous analogy with normally distributed demands. We call the system with the latter cost function an auxiliary (r^λ, q^λ)-system. We also present a similar auxiliary system when the leadtime is deterministic. As we shall show in Section 5, by leveraging the normal distribution, these auxiliary cost functions lend themselves to closed-form *optimal* policy parameters and costs as λ approaches to infinity. There, we shall also show that the optimal behavior for these new systems are equivalent to those of the original systems when λ grows large.

Before proceeding, we introduce the following useful notation and relationships. Define

$$C(z) = h\Phi^1(-z) + p\Phi^1(z), \quad z_* = \Phi^{-1}(p/(p+h)), \tag{32}$$

where $\Phi^1(z) = \int_z^\infty [1-\Phi(x)]\mathsf{d}x$. Note that $C(z)$ is the expected cost of the newsvendor problem with standard normal demand. It can be verified that

$$C'(z) = -p + (p+h)\Phi(z), \quad C''(z) = (p+h)\phi(z), \quad C'(z_*) = 0, \tag{33}$$

where $\phi(\cdot)$ is the standard normal density function. Thus, $C(z)$ is convex and achieves its minimum at z_*. In addition,

$$C_* \equiv C(z_*) = (p+h)\phi(z_*) = C''(z_*). \tag{34}$$

4.1 Random Leadtimes

First, consider random leadtimes. Our goal is to show that, as λ grows large, the expected long-run average system cost of discrete variables $\mathcal{AC}(r^\lambda, q^\lambda)$ in (9) for System-$\mathcal{S}^\lambda$ can be approximated by its continuous analogue

$$\widetilde{\mathcal{AC}}(r^\lambda, q^\lambda) = \frac{1}{q^\lambda}\left\{\lambda K + \gamma^\lambda(q^\lambda)\int_{r^\lambda}^{r^\lambda+q^\lambda} C\left(z^\lambda(x, q^\lambda)\right)\mathsf{d}x\right\}. \tag{35}$$

To see this, note that from (7),

$$\begin{aligned} \mathsf{E}\left[\hat{G}\left(IN^\lambda\right)\right] &= \mathsf{E}\left[\hat{G}\left(r^\lambda + q^\lambda - J^\lambda - qN^\lambda\right)\right] \\ &= \sum_{i=r^\lambda+1}^{r^\lambda+q^\lambda} \mathsf{E}\left[I\left\{J^\lambda = r^\lambda + q^\lambda - i\right\} \times \hat{G}\left(i - q^\lambda \cdot N^\lambda\right)\right]. \end{aligned} \tag{36}$$

In view of (8) and (16), for any i and q^λ,

$$\hat{G}\left(i - q^\lambda \cdot N^\lambda\right) = \gamma^\lambda(q^\lambda) \cdot \hat{G}\left(z^\lambda(i, q^\lambda) - Y^\lambda(q^\lambda)\right). \tag{37}$$

In order to show the approximation for $\mathcal{AC}(r, q)$, we first use (37) to establish the asymptotic expression for each summand in (36), and then obtain the approximation for the sum, as shown in the following lemma. We shall need the following condition on the boundedness of the scaled net inventory level under (r^λ, q^λ)-policy:

Condition 2 *For a sequence of (r^λ, q^λ)-policies,* $\left|\lim_{\lambda\to\infty} z^\lambda(r^\lambda, q^\lambda)\right| < \infty$.

We have the following approximations for the expected long-run average system cost of System-$\mathcal{S}^\lambda$ (see the Appendix for the proof):

Lemma 3 (i) *Assume that the sequence of* (r^λ, q^λ)*-policies satisfies Condition* 1. *For each* i^λ *with* $\left|\lim_{\lambda\to\infty} z^\lambda(i^\lambda, q^\lambda)\right| < \infty$,

$$\lim_{\lambda\to\infty} \mathsf{E}\hat{G}\left(z^\lambda(i^\lambda, q^\lambda) - Y^\lambda(q^\lambda)\right) = \lim_{\lambda\to\infty} C\left(z^\lambda(i^\lambda, q^\lambda)\right). \tag{38}$$

(ii) *If Conditions* 1–2 *hold, then*

$$\lim_{\lambda\to\infty} \Big[\sum_{i=r^\lambda+1}^{r^\lambda+q^\lambda} C\left(z^\lambda(i, q^\lambda)\right) \Big/ \int_{r^\lambda}^{r^\lambda+q^\lambda} C\left(z^\lambda(x, q^\lambda)\right) \mathsf{d}x\Big] = 1. \tag{39}$$

Combining (9), (36)-(37) and Lemma 3, we obtain the following lemma (its proof is given in the Appendix) about the convergence of the expected long-run average system cost (of the original system) to its continuous analogy with normally distributed demands.

Lemma 4 *Under Conditions* 1–2*,*

$$\lim_{\lambda\to\infty} \frac{\mathcal{AC}(r^\lambda, q^\lambda)}{\widetilde{\mathcal{AC}}(r^\lambda, q^\lambda)} = 1. \tag{40}$$

Thus, under Conditions 1–2 with $\lim_{\lambda\to\infty} q^\lambda = \infty$, the (r, q)-system with the long-run average cost given by (35) can be considered as an approximation of the original System-$\mathcal{S}^\lambda$. From now on, we refer to this approximate system as System-$\widetilde{\mathcal{S}}^\lambda$. Its optimal policy is denoted by $(\tilde{r}_*^\lambda, \tilde{q}_*^\lambda)$.

For any given q^λ, let $\tilde{r}_*^\lambda(q^\lambda) = \arg\min_{r^\lambda} \widetilde{\mathcal{AC}}(r^\lambda, q^\lambda)$. By the convexity of $C(\cdot)$ and Lemma 2 of Zheng (1992), we have

$$z^\lambda(\tilde{r}_*^\lambda(q^\lambda), q^\lambda) = z_* - \kappa(q^\lambda), \quad z^\lambda(\tilde{r}_*^\lambda(q^\lambda) + q^\lambda, q^\lambda) = z_* + \overline{\kappa}(q^\lambda), \tag{41}$$

$$C\left(z_* - \kappa(q^\lambda)\right) = C\left(z_* + \overline{\kappa}(q^\lambda)\right), \tag{42}$$

where

$$\kappa(q^\lambda) = \beta^\lambda(q^\lambda) \cdot \alpha(q^\lambda) \;\text{ and }\; \overline{\kappa}(q^\lambda) = \beta^\lambda(q^\lambda) \cdot [1 - \alpha(q^\lambda)] \;\text{ for some } \alpha(q^\lambda) \in [0, 1]. \tag{43}$$

Thus, by (35), we have

$$\min_{r^\lambda, q^\lambda} \widetilde{\mathcal{AC}}(r^\lambda, q^\lambda) = \min_{q^\lambda} \widetilde{\mathcal{AC}}(\tilde{r}_*^\lambda(q^\lambda), q^\lambda) = \min_{q^\lambda} \Big\{\frac{\lambda K}{q^\lambda} + \left(\sigma^\lambda(q^\lambda)\right)^2 \lambda \int_{z_*-\kappa(q^\lambda)}^{z_*+\overline{\kappa}(q^\lambda)} C(x)\mathsf{d}x\Big\}. \tag{44}$$

In the remainder of the paper, we use (44) to analyze system-$\tilde{\mathcal{S}}^\lambda$.

4.2 Constant Leadtimes

Now consider the case of constant leadtimes. To distinguish this case from its random counterpart, we denote everything with a subscript c. In particular, we write System-$\mathcal{S}_c^\lambda$ in place of System-$\mathcal{S}^\lambda$, and denote its optimal policy by $(r_{*c}^\lambda, q_{*c}^\lambda)$.

Let N_c^λ be the number of jobs in steady state in the supply system. We have $D^\lambda = q^\lambda \cdot N_c^\lambda$. Similar to (16), denote

$$z_c^\lambda(i) = \frac{i-\rho^\lambda}{\gamma_c^\lambda}, \quad Y_c^\lambda(q^\lambda) = \beta_c^\lambda(q^\lambda)\Big(N_c^\lambda - \frac{\rho^\lambda}{q^\lambda}\Big) \text{ with } \gamma_c^\lambda = \theta\sqrt{\rho^\lambda} \text{ and } \beta_c^\lambda(q^\lambda) = \frac{q^\lambda}{\gamma_c^\lambda}. \tag{45}$$

Thus, for any i,

$$\mathsf{E}[\hat{G}(i-D^\lambda)] = \gamma_c^\lambda \cdot \mathsf{E}\left[\hat{G}\left(z_c^\lambda(i) - Y_c^\lambda(q^\lambda)\right)\right]. \tag{46}$$

By the central limit theorem for the renewal process (see Theorem 5.1 on p.91, Gut 2009), we have that as $\lambda \to \infty$, $Y_c^\lambda(q^\lambda)$ converges in distribution to a standard normal distribution. Furthermore, $\{Y^\lambda(q^\lambda) : \lambda \geq 1\}$ is uniformly integrable (see Equation (9.1) on p.100, Gut 2009). In order to get the approximation of $\mathcal{AC}_c(r^\lambda, q^\lambda)$, similar to the random leadtime case, we need the following condition.

Condition 3 *For a sequence of* (r^λ, q^λ)*-policies,* $\left|\lim_{\lambda\to\infty} z_c^\lambda(r^\lambda)\right| < \infty$.

Similar to (38)-(39), for the sequence of (r^λ, q^λ)-policies satisfying Condition 3,

$$\lim_{\lambda\to\infty} \mathsf{E}\left[\hat{G}\left(z_c^\lambda(i) - Y_c^\lambda(q^\lambda)\right)\right] = \lim_{\lambda\to\infty} C\left(z_c^\lambda(i)\right), \tag{47}$$

$$\lim_{\lambda\to\infty}\Big(\sum_{i=r^\lambda+1}^{r^\lambda+q^\lambda} C\left(z_c^\lambda(i)\right) \Big/ \int_{r^\lambda}^{r^\lambda+q^\lambda} C\left(z_c^\lambda(x)\right)\mathsf{d}x\Big) = 1. \tag{48}$$

Define

$$\widetilde{\mathcal{AC}}_c(r^\lambda, q^\lambda) = \frac{1}{q^\lambda}\left\{\lambda K + \int_{r^\lambda}^{r^\lambda+q^\lambda} \gamma_c^\lambda \cdot C\left(z_c^\lambda(x)\right)\mathsf{d}x\right\}. \tag{49}$$

By (10) and (46)-(48), similar to Lemma 4, we have

Lemma 5 *Under Conditions* 1 *and* 3*,*

$$\lim_{\lambda\to\infty} \frac{\mathcal{AC}_c(r^\lambda, q^\lambda)}{\widetilde{\mathcal{AC}}_c(r^\lambda, q^\lambda)} = 1. \tag{50}$$

Thus, under Conditions 1–3, the (r, q)-system with the long-run average cost (48) can be considered as an approximation of the original System-$\mathcal{S}_c^\lambda$. From now on, we refer to this approximate system as System-$\widetilde{\mathcal{S}}_c^\lambda$ and denote its optimal policy by $(\tilde{r}_{*c}^\lambda, \tilde{q}_{*c}^\lambda)$.

For any fixed q^λ, denote the optimal reorder point by $\tilde{r}_{*c}^\lambda(q^\lambda) = \arg\min_r \widetilde{\mathcal{AC}}_c(r, q^\lambda)$. Similar to (41)-(43), it can be shown that there exists a unique $\alpha_c(q^\lambda) \in [0, 1]$ such that

$$z_c^\lambda(\tilde{r}_{*c}^\lambda(q^\lambda)) = z_* - \kappa_c(q^\lambda), \qquad z_c^\lambda(\tilde{r}_{*c}^\lambda(q^\lambda) + q^\lambda) = z_* + \overline{\kappa}_c(q^\lambda), \tag{51}$$

$$C(z_* - \kappa_c(q^\lambda)) = C(z_* + \overline{\kappa}_c(q^\lambda)), \tag{52}$$

where $\kappa_c(q^\lambda) = \alpha_c(q^\lambda) \cdot \beta_c^\lambda$ and $\overline{\kappa}_c(q^\lambda) = [1 - \alpha_c(q^\lambda)] \cdot \beta_c^\lambda$. Furthermore, because γ_c^λ is independent of the order quantity (unlike γ^λ for the random leadtime case), by Lemma 6 of Zheng (1992), the optimal order quantity is the solution of

$$\lambda K \quad = \quad q^\lambda \cdot \gamma_c^\lambda \cdot C\left(z_c^\lambda(\tilde{r}_{*c}^\lambda(q^\lambda))\right) - (\gamma_c^\lambda)^2 \int_{z_* - \kappa_c(q^\lambda)}^{z_* + \overline{\kappa}_c(q^\lambda)} C(x)\mathsf{d}x. \tag{53}$$

In other words, the optimal policy satisfies

$$\frac{\lambda K}{\gamma_c^\lambda \tilde{q}_{*c}^\lambda} = C(z_* - \kappa_c(\tilde{q}_{*c}^\lambda)) - \frac{\gamma_c^\lambda}{\tilde{q}_{*c}^\lambda} \int_{z_* - \kappa_c(\tilde{q}_{*c}^\lambda)}^{z_* + \overline{\kappa}_c(\tilde{q}_{*c}^\lambda)} C(x)\mathsf{d}x. \tag{54}$$

In the remainder of the paper, we shall use (54) to analyze System-$\tilde{\mathcal{S}}_c^\lambda$.

Comparing (44) with (49), we can see that $\gamma^\lambda(q^\lambda)$ in the approximate cost under random leadtime depends on the decision variable q^λ, whereas γ_c^λ under constant leadtime does not. This difference yields different first-order-conditions for the optimization problems in System-$\tilde{\mathcal{S}}$ and System-$\tilde{\mathcal{S}}_c^\lambda$. More importantly, the latter is not a special case of the former. As a result, the subsequent analyses of the optimal solutions of these two system in the next section will be different.

5 Asymptotic Behavior of the Optimal Policy and Cost

In this section we analyze the asymptotic behavior of the optimal policy and cost.

5.1 Main Results

We first need the following lemma to describe our main results; its proof can be found in the Appendix.

Lemma 6 *If $K > 0$, then there exists a unique solution (τ, α) with $\tau \in (0, \infty)$ and $\alpha \in [0, 1]$ to equations*

$$\eta K = \tau \times C\left(z_* - \alpha\tau\eta\right) - \frac{1}{\eta}\int_{z_*-\alpha\tau\eta}^{z_*+(1-\alpha)\tau\eta} C(y)\mathsf{d}y, \tag{55}$$

$$C\left(z_* - \alpha\tau\eta\right) = C\left(z_* + (1-\alpha)\tau\eta\right). \tag{56}$$

Now we present our main results.

Theorem 2 *Let τ and α be defined as in* (55)-(56)*. If the replenishment leadtime is random, then*

$$\text{(A.i)}\quad q_*^\lambda = \begin{cases} \left(\frac{2K}{C_*\sqrt{\nu}}\right)^{2/3}\lambda^{1/3} + o\left(\lambda^{1/3}\right), & \text{if } K > 0, \\ 1, & \text{if } K = 0; \end{cases}$$

$$\text{(A.ii)}\quad r_*^\lambda = \begin{cases} \lambda/\mu + z_* \cdot \left(\frac{2K\nu}{C_*}\right)^{1/3} \cdot \lambda^{2/3} + o\left(\lambda^{2/3}\right), & \text{if } K > 0, \\ \lambda/\mu + z_*\sqrt{(1-\theta^2)\nu + \frac{\theta^2}{\mu}} \cdot \sqrt{\lambda} + o(\sqrt{\lambda}), & \text{if } K = 0; \end{cases}$$

$$\text{(A.iii)}\quad \mathcal{AC}(r_*^\lambda, q_*^\lambda) = \begin{cases} 3\left(\frac{K\nu C_*^2}{4}\right)^{1/3}\lambda^{2/3} + o\left(\lambda^{2/3}\right), & \text{if } K > 0, \\ C_*\sqrt{(1-\theta^2)\nu + \frac{\theta^2}{\mu}} \cdot \sqrt{\lambda} + o(\sqrt{\lambda}), & \text{if } K = 0. \end{cases}$$

If the replenishment leadtime is a constant $1/\mu$, then

$$\text{(B.i)}\quad q_{*c}^\lambda = \begin{cases} \tau\sqrt{\lambda} + o(\sqrt{\lambda}), & \text{if } K > 0, \\ 1, & \text{if } K = 0; \end{cases}$$

$$\text{(B.ii)}\quad r_{*c}^\lambda = \begin{cases} \lambda/\mu + z_*\theta\sqrt{\lambda/\mu} - \alpha\tau\sqrt{\lambda} + o(\sqrt{\lambda}), & \text{if } K > 0, \\ \lambda/\mu + z_*\theta\sqrt{\lambda/\mu} + o(\sqrt{\lambda}), & \text{if } K = 0; \end{cases}$$

$$\text{(B.iii)}\quad \mathcal{AC}_c(r_{*c}^\lambda, q_{*c}^\lambda) = \begin{cases} C\left(z_* - \alpha\tau\sqrt{\mu}/\theta\right)\theta\sqrt{\lambda/\mu} + o(\sqrt{\lambda}), & \text{if } K > 0, \\ C_*\theta\sqrt{\lambda/\mu} + o(\sqrt{\lambda}), & \text{if } K = 0. \end{cases}$$

Remark 1 When $K = 0$ and the demand process is Poisson, (A.ii) and (A.iii) give

$$r_*^\lambda = \lambda/\mu + z_*\sqrt{\lambda/\mu} + o(\sqrt{\lambda}), \qquad \mathcal{AC}(r_*^\lambda, q_*^\lambda) = C_* \cdot \sqrt{\lambda/\mu} + o(\sqrt{\lambda}).$$

These agree with the standard approximate formula $\lambda\mathsf{E}[L] + z_*\sqrt{\lambda\mathsf{E}[L]}$ for the optimal base-stock level (based on the normal approximation for the outstanding orders) and the resulting approximate optimal cost; see Zipkin (2000, Chapter 7). On the other hand, (B.ii) and (B.iii) give

$$r_{*c}^\lambda = \lambda/\mu + z_*\theta\sqrt{\lambda/\mu} + o(\sqrt{\lambda}), \qquad \mathcal{AC}_c(r_{*c}^\lambda, q_{*c}^\lambda) = C_* \cdot \theta\sqrt{\lambda/\mu} + o(\sqrt{\lambda}).$$

Here, θ measures inter-demand variability, which equals 1 for Poisson demand. Thus, our asymptotic analysis reveals new insights on the effect of demand variability.

Remark 2 From (A.i) and (B.i), it is striking that the well-known square-root relationship between the optimal order quantity and the demand rate holds only for the extreme case of constant leadtimes. At this extreme, the sequential and parallel processing environments converge. Under general i.i.d. leadtimes, the square-root relationship is replaced by the cube root. Thus, as demand rate increases, the optimal order quantity grows more slowly in a stochastic parallel processing environment than in a sequential processing environment. In addition, the leadtime variability contracts this relationship by a factor of $\nu^{1/3}$, while the fixed cost amplifies this relationship by a factor of $K^{2/3}$.

Remark 3 With $K > 0$ and i.i.d. stochastic leadtimes, (A.i) shows that in the asymptotic regime, the optimal order quantity q_*^λ increases as $K^{2/3}$, which is faster than the EOQ formula that is proportional to $K^{1/2}$. Moreover, (A.ii) and (A.iii) indicate that both the asymptotically optimal reorder point r_*^λ and cost increase in K as well as in leadtime variability (measured by ν). Interestingly, both the safety stock (i.e., the second term of r_*^λ) and the optimal cost increase in the demand rate λ faster than the well-known square-root law.

Remark 4 When the leadtime is exogenous and sequential, Gallego (1998) derives bounds on q_* which depend on the variance of the leadtime that is of higher order than $\sqrt{\lambda}$. For the special case of a constant leadtime, which is applicable to both his and our settings, our result in Theorem 2 (B.i) gives a more accurate estimate for q_*^λ than his bounds. Moreover, the gap between his lower and upper bounds widens as λ increases.

5.2 Analysis: Random Leadtimes

In this subsection we prove Part A of Theorem 2. Here is the basic idea, which contains three steps. In Step 1, we show that the optimal policy $(\tilde{r}_*^\lambda, \tilde{q}_*^\lambda)$ of the auxiliary System-$\tilde{\mathcal{S}}^\lambda$ satisfies the properties of Part A of Theorem 2; see Proposition 1. Hence, property (A.iii) for the original System-$\mathcal{S}^\lambda$ will be established if we can show

$$\lim_{\lambda\to\infty} \frac{\mathcal{AC}(r_*^\lambda, q_*^\lambda)}{\widetilde{\mathcal{AC}}(\tilde{r}_*^\lambda, \tilde{q}_*^\lambda)} = 1. \tag{57}$$

Because

$$\frac{\mathcal{AC}(r_*^\lambda, q_*^\lambda)}{\widetilde{\mathcal{AC}}(\tilde{r}_*^\lambda, \tilde{q}_*^\lambda)} = \frac{\mathcal{AC}(r_*^\lambda, q_*^\lambda)}{\widetilde{\mathcal{AC}}(r_*^\lambda, q_*^\lambda)} \cdot \frac{\widetilde{\mathcal{AC}}(r_*^\lambda, q_*^\lambda)}{\widetilde{\mathcal{AC}}(\tilde{r}_*^\lambda, \tilde{q}_*^\lambda)}, \tag{58}$$

what remain to be shown is

$$\lim_{\lambda\to\infty}\frac{\mathcal{AC}(r_*^\lambda,q_*^\lambda)}{\widetilde{\mathcal{AC}}(r_*^\lambda,q_*^\lambda)}=1 \tag{59}$$

and

$$\lim_{\lambda\to\infty}\frac{\widetilde{\mathcal{AC}}(r_*^\lambda,q_*^\lambda)}{\widetilde{\mathcal{AC}}(\tilde{r}_*^\lambda,\tilde{q}_*^\lambda)}=1. \tag{60}$$

In addition, given (60), if we can show the uniqueness for the first and second leading terms of the asymptotic optimal reorder point ($\tilde{r}_*^\lambda$), and the uniqueness for the leading term of the asymptotic optimal ordering quantity ($\tilde{q}_*^\lambda$) and cost of System-$\tilde{\mathcal{S}}^\lambda$, then (A.i) and (A.ii) will hold for System-$\mathcal{S}^\lambda$. Step 2 proves these uniqueness properties (see Proposition 2). Step 3 establishes (59) and (60); see Propositions 3 and 4.

We now start at Step 1 – to show System-$\tilde{\mathcal{S}}^\lambda$ possesses the properties (A.i)-(A.iii). Consider $K>0$. We would like to work on the optimization problem (44) by the first order condition. To do so, we first need the following result about the differentiability on our objective function (see the Appendix for a proof).

Lemma 7 *$\widetilde{\mathcal{AC}}(\tilde{r}_*^\lambda(q^\lambda),q^\lambda)$ is differentiable with respect to q^λ.*

Now, using (42) and the first order condition on $\widetilde{\mathcal{AC}}(\tilde{r}_*^\lambda(q^\lambda),q^\lambda)$ given by (44), we know that the optimal solution $\tilde{q}_*^\lambda$ satisfies

$$\frac{\lambda K}{q^2}=\frac{\mathsf{d}(\sigma^\lambda(q))^2}{\mathsf{d}q}\lambda\int_{z_*-\kappa(q)}^{z_*+\overline{\kappa}(q)}C(x)\mathsf{d}x+(\sigma^\lambda(q))^2\lambda\cdot C\left(z_*-\kappa(q)\right)\cdot\frac{\mathsf{d}}{\mathsf{d}q}\sqrt{\frac{q}{\lambda(\sigma^\lambda(q))^2}}. \tag{61}$$

In view of (14),

$$\frac{\mathsf{d}(\sigma^\lambda(q))^2}{\mathsf{d}q}=-\frac{\theta^2}{q^2}\mathsf{E}[L_{(2)}]\ \text{ and }\ \frac{\mathsf{d}}{\mathsf{d}q}\sqrt{\frac{q}{\lambda(\sigma^\lambda(q))^2}}=\frac{\nu q+2\theta^2\mathsf{E}[L_{(2)}]}{2\sqrt{\lambda}\sqrt{(\nu q+\theta^2\mathsf{E}[L_{(2)}])^3}}.$$

Plugging these into (61) yields

$$K=-\theta^2\mathsf{E}[L_{(2)}]\int_{z_*-\kappa(q)}^{z_*+\overline{\kappa}(q)}C(x)\mathsf{d}x+\frac{q}{2\sqrt{\lambda}}\cdot\frac{\nu q+2\theta^2\mathsf{E}[L_{(2)}]}{\sqrt{\nu q+\theta^2\mathsf{E}[L_{(2)}]}}\cdot C\left(z_*-\kappa(q)\right). \tag{62}$$

Its solution gives $\tilde{q}_*^\lambda$. However, it is difficult to solve this equation directly, so we resort to its Taylor expansion for an approximate solution. To validate the expansion, we need the following lemma; its proof is provided in the Appendix.

Lemma 8 *The sequence of optimal order sizes $\tilde{q}_*^\lambda$ for system-$\tilde{\mathcal{S}}^\lambda$ with $K > 0$ satisfies Condition 1 and $\lim_{\lambda\to\infty} \tilde{q}_*^\lambda = \infty$. Moreover, $\lim_{\lambda\to\infty} \kappa(\tilde{q}_*^\lambda) = \lim_{\lambda\to\infty} \overline{\kappa}(\tilde{q}_*^\lambda) = 0$.*

With the help of Lemma 8, we can show the following asymptotic behavior of system-$\widetilde{\mathcal{S}}^\lambda$.

Proposition 1 *The optimal policy $(\tilde{r}_*^\lambda, \tilde{q}_*^\lambda)$ and cost $\widetilde{\mathcal{AC}}(\tilde{r}_*^\lambda, \tilde{q}_*^\lambda)$ for System-$\tilde{\mathcal{S}}^\lambda$ satisfy Theorem 2 (A.i)-(A.iii), respectively.*

Proof : First, assume $K > 0$. By Lemma 8 and the Taylor expansion, and recalling (33)-(34), we have

$$\begin{aligned}
&\int_{z_*-\kappa(\tilde{q}_*^\lambda)}^{z_*+\overline{\kappa}(\tilde{q}_*^\lambda)} C(y)\mathsf{d}y \\
&= \int_{z_*-\kappa(\tilde{q}_*^\lambda)}^{0} C(y)\mathsf{d}y + \int_0^{z_*+\overline{\kappa}(\tilde{q}_*^\lambda)} C(y)\mathsf{d}y \\
&= \int_{z_*}^{0} C(y)\mathsf{d}y + C_*\kappa(\tilde{q}_*^\lambda) - \frac{1}{2!}C'(z_*)\Big(\kappa(\tilde{q}_*^\lambda)\Big)^2 + \frac{1}{3!}C''(z_*)\left(\kappa(\tilde{q}_*^\lambda)\right)^3 + O\Big(\Big(\beta^\lambda(\tilde{q}_*^\lambda)\Big)^4\Big) \\
&\quad + \int_0^{z_*} C(y)\mathsf{d}y + C_*\overline{\kappa}(\tilde{q}_*^\lambda) + \frac{1}{2!}C'(z_*)\left(\overline{\kappa}(\tilde{q}_*^\lambda)\right)^2 + \frac{1}{3!}C''(z_*)\left(\overline{\kappa}(\tilde{q}_*^\lambda)\right)^3 + O\Big(\Big(\beta^\lambda(\tilde{q}_*^\lambda)\Big)^4\Big) \\
&= C_*\beta^\lambda(\tilde{q}_*^\lambda) + \frac{C_*}{3!}\left(\beta^\lambda(\tilde{q}_*^\lambda)\right)^3\left(1 - 3\alpha(\tilde{q}_*^\lambda) + 3\alpha^2(\tilde{q}_*^\lambda)\right) + O\Big(\Big(\beta^\lambda(\tilde{q}_*^\lambda)\Big)^4\Big). \qquad (63)
\end{aligned}$$

By again the Taylor expansion,

$$\begin{aligned}
C(z_* - \kappa(\tilde{q}_*^\lambda)) &= C_* - C'(z_*)\kappa(\tilde{q}_*^\lambda) + \frac{1}{2!}C''(z_*)\left(\kappa(\tilde{q}_*^\lambda)\right)^2 + O\Big(\Big(\beta^\lambda(\tilde{q}_*^\lambda)\Big)^3\Big) \\
&= C_* + \frac{1}{2!}C_* \cdot \left(\kappa(\tilde{q}_*^\lambda)\right)^2 + O\Big(\Big(\beta^\lambda(\tilde{q}_*^\lambda)\Big)^3\Big). \qquad (64)
\end{aligned}$$

Note that

$$\frac{\nu q + 2\theta^2\mathsf{E}[L_{(2)}]}{\sqrt{\nu q + \theta^2\mathsf{E}[L_{(2)}]}} = \sqrt{\nu q + \theta^2\mathsf{E}[L_{(2)}]} + \frac{\theta^2\mathsf{E}[L_{(2)}]}{\sqrt{\nu q + \theta^2\mathsf{E}[L_{(2)}]}}. \qquad (65)$$

It follows from Lemma 8 and (62)-(65) that

$$\tilde{q}_*^\lambda = \left(\frac{2K}{C_*\sqrt{\nu}}\right)^{2/3} \cdot \lambda^{1/3} + o\left(\lambda^{1/3}\right), \qquad (66)$$

which is (A.i) for $K > 0$.

Now we examine $\tilde{r}_*^\lambda$. By the Taylor expansion of both sides of (42) (expanding to the second moment), we have

$$\begin{aligned}
&C_* - C'(z_*)\kappa(\tilde{q}_*^\lambda) + \frac{1}{2}C''(z_*)\left(\kappa(\tilde{q}_*^\lambda)\right)^2 \\
&\quad = C_* + C'(z_*)\overline{\kappa}(\tilde{q}_*^\lambda) + \frac{1}{2}C''(z_*)\left(\overline{\kappa}(\tilde{q}_*^\lambda)\right)^2 + O\left(\left(\beta^\lambda(\tilde{q}_*^\lambda)\right)^3\right).
\end{aligned}$$

Applying (33)-(34) yields $(1/2)\left(1 - 2\alpha(\tilde{q}_*^\lambda)\right) \cdot C_* = O\left(\beta^\lambda(\tilde{q}_*^\lambda)\right)$. Because C_* is a positive constant, we know that

$$\alpha(\tilde{q}_*^\lambda) = \frac{1}{2} + O\left(\beta^\lambda(\tilde{q}_*^\lambda)\right). \tag{67}$$

Thus, by (16), (41) and (43),

$$\tilde{r}_*^\lambda = \rho^\lambda + z_* \cdot \left(\frac{2K\nu}{C_*}\right)^{1/3} \cdot \lambda^{2/3} + o\left(\lambda^{2/3}\right). \tag{68}$$

This is (A.ii) for $K > 0$.

For the optimal cost of System-$\tilde{\mathcal{S}}^\lambda$, following (44), (63), and (66)-(68),

$$\begin{aligned} \widetilde{\mathcal{AC}}(\tilde{r}_*^\lambda, \tilde{q}_*^\lambda) &= \frac{\lambda K}{\tilde{q}_*^\lambda} + \left(\sigma^\lambda(\tilde{q}_*^\lambda)\right)^2 \lambda \int_{z_* - \kappa(\tilde{q}_*^\lambda)}^{z_* + \overline{\kappa}(\tilde{q}_*^\lambda)} C(y) \mathsf{d}y \\ &= \lambda K \Big/ \Big[\Big(\frac{2K}{C_*\sqrt{\nu}}\Big)^{2/3} \cdot \lambda^{1/3} + o\left(\lambda^{1/3}\right) \Big] \\ &\quad + \left(\sigma^\lambda(\tilde{q}_*^\lambda)\right)^2 \lambda \left[C_* \beta^\lambda(\tilde{q}_*^\lambda) + \frac{C_*}{4!} \left(\beta^\lambda(\tilde{q}_*^\lambda)\right)^3 + O\left(\left(\beta^\lambda(\tilde{q}_*^\lambda)\right)^4\right) \right] \\ &= 3\Big(\frac{K\nu C_*^2}{4}\Big)^{1/3} \lambda^{2/3} + o\left(\lambda^{2/3}\right). \end{aligned} \tag{69}$$

This is (A.iii) for $K > 0$.

When $K = 0$, $\tilde{q}_*^\lambda = 1$ is our assumption; (A.ii) and (A.iii) are given by (35). ■

Next, we proceed to Step 2 – to show the uniqueness of optimal policy $(\tilde{r}_*^\lambda, \tilde{q}_*^\lambda)$ for System-$\tilde{\mathcal{S}}^\lambda$ described at the beginning of this subsection (see right after (60)). That is, in view of (60), for any $(\tilde{r}^\lambda, \tilde{q}^\lambda)$ satisfying $\lim_{\lambda\to\infty} \widetilde{\mathcal{AC}}(\tilde{r}^\lambda, \tilde{q}^\lambda)/\widetilde{\mathcal{AC}}(\tilde{r}_*^\lambda, \tilde{q}_*^\lambda) = 1$, $\tilde{r}^\lambda$ and $\tilde{q}^\lambda$ should have the same order of $\tilde{r}_*^\lambda$ and $\tilde{q}_*^\lambda$, respectively. As each of $(\tilde{r}^\lambda, \tilde{q}^\lambda)$, $\widetilde{\mathcal{AC}}(\tilde{r}^\lambda, \tilde{q}^\lambda)$, $(\tilde{r}_*^\lambda, \tilde{q}_*^\lambda)$ and $\widetilde{\mathcal{AC}}(\tilde{r}_*^\lambda, \tilde{q}_*^\lambda)$ will go to infinity when λ grows large, we need to use alternative measures to characterize the uniqueness. To see this, note that from Proposition 1, $\tilde{q}_*^\lambda$ and $\widetilde{\mathcal{AC}}(\tilde{r}_*^\lambda, \tilde{q}_*^\lambda)$ have one dominant term given by the order of $\lambda^{1/3}$ and $\lambda^{2/3}$, respectively. Thus their uniqueness can be characterized directly by their corresponding ratios $\tilde{q}^\lambda/\tilde{q}_*^\lambda$ and $\widetilde{\mathcal{AC}}(\tilde{r}^\lambda, \tilde{q}^\lambda)/\widetilde{\mathcal{AC}}(\tilde{r}_*^\lambda, \tilde{q}_*^\lambda)$. However, $\tilde{r}_*^\lambda$ has two dominant terms, λ and $\lambda^{2/3}$. The ratio $\tilde{r}^\lambda/\tilde{r}_*^\lambda$ cannot characterize the uniqueness about the term $\lambda^{2/3}$ when λ grows large. For this reason, we consider the scaled net inventory level $z^\lambda(\tilde{r}^\lambda, \tilde{q}^\lambda)$ instead of $\tilde{r}^\lambda/\tilde{r}_*^\lambda$. More specifically, we have

Proposition 2 *Let* $\Delta(\lambda) = \tilde{q}^\lambda/\tilde{q}_*^\lambda$.

(i) *If* $\overline{\lim}_{\lambda\to\infty} \Delta(\lambda) \neq 1$ *or* $\underline{\lim}_{\lambda\to\infty} \Delta(\lambda) \neq 1$ *holds, then*

$$\overline{\lim_{\lambda\to\infty}} \frac{\min_{r^\lambda} \widetilde{\mathcal{AC}}(r^\lambda, \tilde{q}^\lambda)}{\widetilde{\mathcal{AC}}(\tilde{r}_*^\lambda, \tilde{q}_*^\lambda)} > 1.$$

(ii) *Assume that* $\lim_{\lambda\to\infty} \Delta(\lambda) = 1$ *and* $\overline{\lim}_{\lambda\to\infty} |z^\lambda(\tilde{r}^\lambda, \tilde{q}^\lambda)| < \infty$*. If* $\overline{\lim}_{\lambda\to\infty} z^\lambda(\tilde{r}^\lambda, \tilde{q}^\lambda) \neq z_*$ *or* $\underline{\lim}_{\lambda\to\infty} z^\lambda(\tilde{r}^\lambda, \tilde{q}^\lambda) \neq z_*$ *holds, then*

$$\overline{\lim_{\lambda\to\infty}} \frac{\widetilde{\mathcal{AC}}(\tilde{r}^\lambda, \tilde{q}^\lambda)}{\widetilde{\mathcal{AC}}(\tilde{r}_*^\lambda, \tilde{q}_*^\lambda)} > 1.$$

Proof : First by (44),

$$\min_{r^\lambda} \widetilde{\mathcal{AC}}(r^\lambda, \tilde{q}^\lambda) = \frac{\lambda K}{\tilde{q}^\lambda} + \left(\sigma^\lambda(\tilde{q}^\lambda)\right)^2 \lambda \int_{z_*-\kappa(\tilde{q}^\lambda)}^{z_*+\overline{\kappa}(\tilde{q}^\lambda)} C(x)\mathsf{d}x. \tag{70}$$

If $\overline{\lim}_{\lambda\to\infty} \tilde{q}^\lambda/\lambda > 0$, then, by (43), we have

$$\overline{\lim_{\lambda\to\infty}} \int_{z_*-\kappa(\tilde{q}^\lambda)}^{z_*+\overline{\kappa}(\tilde{q}^\lambda)} C(x)\mathsf{d}x > 0.$$

This together with (70) gives that

$$\overline{\lim_{\lambda\to\infty}} \frac{1}{\lambda} \left\{ \min_{r^\lambda} \widetilde{\mathcal{AC}}(r^\lambda, \tilde{q}^\lambda) \right\} > 0.$$

This, by Proposition 1, implies that

$$\underline{\lim}_{\lambda\to\infty} \frac{\min_{r^\lambda} \widetilde{\mathcal{AC}}(r^\lambda, \tilde{q}^\lambda)}{\widetilde{\mathcal{AC}}(\tilde{r}_*^\lambda, \tilde{q}_*^\lambda)} = \infty.$$

Hence, to prove the proposition, it suffices to consider $\overline{\lim}_{\lambda\to\infty} \tilde{q}^\lambda/\lambda = 0$. Under this condition, by the Taylor expansion given by (63),

$$\int_{z_*-\kappa(\tilde{q}^\lambda)}^{z_*+\overline{\kappa}(\tilde{q}^\lambda)} C(x)\mathsf{d}x = C_*\beta^\lambda(\tilde{q}^\lambda) + \frac{C_*}{3!}\left(\beta^\lambda(\tilde{q}^\lambda)\right)^3 \left(1 - 3\alpha(\tilde{q}^\lambda) + 3\alpha^2(\tilde{q}^\lambda)\right) + O\left(\left(\beta^\lambda(\tilde{q}^\lambda)\right)^4\right).$$

Hence,

$$\begin{aligned}
&\frac{\lambda K}{\tilde{q}^\lambda} + \left(\sigma^\lambda(\tilde{q}^\lambda)\right)^2 \lambda \int_{z_*-\kappa(\tilde{q}^\lambda)}^{z_*+\overline{\kappa}(\tilde{q}^\lambda)} C(x)\mathsf{d}x \\
&= \frac{\lambda K}{\Delta(\lambda)\tilde{q}_*^\lambda} + \left(\sigma^\lambda(\tilde{q}^\lambda)\right)^2 \lambda \Big[C_*\beta^\lambda(\tilde{q}^\lambda) + \frac{C_*}{3!}\left(\beta^\lambda(\tilde{q}^\lambda)\right)^3 \left(1 - 3\alpha(\tilde{q}^\lambda) + 3\alpha^2(\tilde{q}^\lambda)\right) \\
&\qquad + O\left(\left(\beta^\lambda(\tilde{q}^\lambda)\right)^4\right)\Big] \\
&= \Big[\Big(\frac{K\nu C_*^2}{4}\Big)^{1/3} \cdot \frac{1}{\Delta(\lambda)} + \left(2K\nu C_*^2\right)^{1/3}\sqrt{\Delta(\lambda)}\Big] \cdot \lambda^{2/3} + o\left(\lambda^{2/3}\right).
\end{aligned} \tag{71}$$

Let

$$U(x) = \left[\left(\frac{K\nu C_*^2}{4}\right)^{1/3} \cdot \frac{1}{x} + \left(2K\nu C_*^2\right)^{1/3}\sqrt{x}\right]$$

It is direct to verify that $-U(x)$ is unimodal, and $\arg\min_x U(x) = 1$. Therefore, (71) implies part (i) of the proposition.

Next, consider part (ii). Note, by (35), that

$$\widetilde{\mathcal{AC}}(\tilde{r}^\lambda, \tilde{q}^\lambda) \quad = \quad \frac{\lambda K}{\tilde{q}^\lambda} + \left(\sigma^\lambda(\tilde{q}^\lambda)\right)^2 \lambda \int_{z^\lambda(\tilde{r}^\lambda,\tilde{q}^\lambda)}^{z^\lambda(\tilde{r}^\lambda,\tilde{q}^\lambda)+\beta^\lambda(\tilde{q}^\lambda)} C(y)\mathsf{d}y. \tag{72}$$

If $\overline{\lim}_{\lambda\to\infty} z^\lambda(\tilde{r}^\lambda, \tilde{q}^\lambda) \neq z_*$ or $\underline{\lim}_{\lambda\to\infty} z^\lambda(\tilde{r}^\lambda, \tilde{q}^\lambda) \neq z_*$, then there exists a subsequence, again writing as λ, such that

$$\lim_{\lambda\to\infty} z^\lambda(\tilde{r}^\lambda, \tilde{q}^\lambda) = b \neq z_*. \tag{73}$$

Now making the Taylor expansion (expanding to the second moment) for the last term in (72), we obtain

$$\left(\sigma^\lambda(\tilde{q}^\lambda)\right)^2 \lambda \int_{z^\lambda(\tilde{r}^\lambda,\tilde{q}^\lambda)}^{z^\lambda(\tilde{r}^\lambda,\tilde{q}^\lambda)+\beta^\lambda(\tilde{q}^\lambda)} C(y)\mathsf{d}y = \sigma^\lambda(\tilde{q}^\lambda) \cdot C(z^\lambda(\tilde{r}^\lambda, \tilde{q}^\lambda))\sqrt{\lambda\tilde{q}^\lambda} + o\left(\lambda^{2/3}\right). \tag{74}$$

By the definition of z_* and (73), we know that $\lim_{\lambda\to\infty} C(z^\lambda(\tilde{r}^\lambda, \tilde{q}^\lambda)) = C(b) > C(z_*)$. This together with (72) and (74) yields part (ii). ■

Finally, we perform Step 3: to show (59) and (60). To prove (59), by Lemma 4, it is sufficient to verify that the optimal policy $(r_*^\lambda, q_*^\lambda)$ satisfies Conditions 1–2. To this end, we first establish Condition 1 and the optimal order quantity q_*^λ will become large when the demand rate λ grows large.

Proposition 3 *The sequence of optimal order sizes q_*^λ for system-$\mathcal{S}^\lambda$ with $K > 0$ satisfies Condition* 1 *and* $\lim_{\lambda\to\infty} q_*^\lambda = \infty$.

Proof : Suppose contrariwise that the proposition is not true. Then there exists a subsequence $\{\lambda_k : k \geq 1\}$ such that

$$\lim_{k\to\infty} q_*^{\lambda_k} < \infty \quad \text{or} \quad \lim_{k\to\infty} \frac{q_*^{\lambda_k}}{\lambda_k} > 0. \tag{75}$$

To simplify notation, we write the sequence as λ (In the remainder of the paper, for the same reason, the subsequences will be always written as λ). By Lemma 1, we know that given q^λ,

$q^\lambda \| \left(J^\lambda + q^\lambda \cdot N^\lambda\right)$ is uniformly distributed on $\mathcal{Q}$. Here, again, "$\|$" is the modulo operator. Now let

$$\begin{aligned}
\Delta_1^\lambda &= \Big\{0, \cdots, \lfloor \frac{q_*^\lambda}{4} \rfloor - 1\Big\}, \quad \Delta_2^\lambda = \Big\{\lfloor \frac{q_*^\lambda}{4} \rfloor, \cdots, 2\lfloor \frac{q_*^\lambda}{4} \rfloor - 1\Big\}, \\
\Delta_3^\lambda &= \Big\{2\lfloor \frac{q_*^\lambda}{4} \rfloor, \cdots, 3\lfloor \frac{q_*^\lambda}{4} \rfloor - 1\Big\}, \quad \Delta_4^\lambda = \Big\{3\lfloor \frac{q_*^\lambda}{4} \rfloor, \cdots, q_*^\lambda - 1\Big\}.
\end{aligned}$$

When $(q_*^\lambda \| r_*^\lambda) \in \Delta_1^\lambda$, we have

$$\left| r_*^\lambda + q_*^\lambda - J^\lambda - q_*^\lambda \cdot N^\lambda \right| \times I\{J^\lambda \in \Delta_3^\lambda\} \geq \lfloor \frac{q_*^\lambda}{4} \rfloor - 1.$$

Hence, if $(q_*^\lambda \| r_*^\lambda) \in \Delta_1^\lambda$ and the second inequality in (75) holds, then

$$\begin{aligned}
\overline{\lim_{\lambda\to\infty}} \frac{\mathcal{AC}(r_*^\lambda, q_*^\lambda)}{\lambda} &\geq \overline{\lim_{\lambda\to\infty}} \frac{1}{\lambda} \mathsf{E}\left[\hat{G}\left(r_*^\lambda + q_*^\lambda - J^\lambda - q_*^\lambda \cdot N^\lambda\right)\right] \quad \text{(by (9))} \\
&\geq \overline{\lim_{\lambda\to\infty}} \frac{1}{\lambda} \mathsf{E}\left[\hat{G}\left(r_*^\lambda + q_*^\lambda - J^\lambda - q_*^\lambda \cdot N^\lambda\right) \times I\{J^\lambda \in \Delta_3^\lambda\}\right] \\
&\geq \overline{\lim_{\lambda\to\infty}} \frac{1}{\lambda} \mathsf{E}\Big[\min\{p,h\} \times \left(\lfloor \frac{q_*^\lambda}{4} \rfloor - 1\right) \times I\{J^\lambda \in \Delta_3^\lambda\}\Big] \\
&\geq \lim_{\lambda\to\infty} \frac{\min\{p,h\}}{\lambda} \times \frac{1}{4} \times \left(\lfloor \frac{q_*^\lambda}{4} \rfloor - 1\right) \\
&> 0. \qquad (76)
\end{aligned}$$

Similarly, we can show that for $(q_*^\lambda \| r_*^\lambda) \in \Delta_i^\lambda$ $(i = 2, 3, 4)$, (76) still holds if the second inequality in (75) holds.

If the first inequality in (75) holds, then by $\mathsf{E}[\hat{G}(IN)] \geq 0$ and (9),

$$\overline{\lim_{\lambda\to\infty}} \frac{\mathcal{AC}(r_*^\lambda, q_*^\lambda)}{\lambda} \geq \overline{\lim_{\lambda\to\infty}} \frac{K}{q_*^\lambda} > 0. \qquad (77)$$

By the definition of $z^\lambda(i, q^\lambda)$ (see (16)), for $i = \lfloor \rho^\lambda \rfloor + 1, \cdots, \lfloor \rho^\lambda \rfloor + \lfloor \sqrt{\lambda} \rfloor$, $\lim_{\lambda\to\infty} z^\lambda(i, \lfloor \sqrt{\lambda} \rfloor) = 0$. So when policy $(r^\lambda, q^\lambda) = (\lfloor \rho^\lambda \rfloor, \lfloor \sqrt{\lambda} \rfloor)$ is implemented, by (9) and (36)-(38), we have

$$\begin{aligned}
\lim_{\lambda\to\infty} \frac{\mathcal{AC}(\lfloor \rho^\lambda \rfloor, \lfloor \sqrt{\lambda} \rfloor)}{\lambda} &= \lim_{\lambda\to\infty} \frac{1}{\lambda} \left\{ \frac{\lambda K}{\lfloor \sqrt{\lambda} \rfloor} + \frac{\sigma^\lambda(\lfloor \sqrt{\lambda} \rfloor)\sqrt{\lambda \lfloor \sqrt{\lambda} \rfloor}}{\lfloor \sqrt{\lambda} \rfloor} \sum_{i=\lfloor \rho^\lambda \rfloor+1}^{\lfloor \rho^\lambda \rfloor + \lfloor \sqrt{\lambda} \rfloor} C\left(z^\lambda(i, \lfloor \sqrt{\lambda} \rfloor)\right) \right\} \\
&\leq \lim_{\lambda\to\infty} \frac{\sigma^\lambda(\lfloor \sqrt{\lambda} \rfloor)\lfloor \sqrt{\lambda} \rfloor}{\sqrt{\lambda \lfloor \sqrt{\lambda} \rfloor}} \cdot \max_{\lfloor \rho^\lambda \rfloor + 1 \leq i \leq \lfloor \rho^\lambda \rfloor + \lfloor \sqrt{\lambda} \rfloor} C\left(z^\lambda(i, \lfloor \sqrt{\lambda} \rfloor)\right) \\
&= 0.
\end{aligned}$$

So in view of (76)-(77), when (75) holds,

$$\lim_{\lambda\to\infty} \frac{\mathcal{AC}(r_*^\lambda, q_*^\lambda)}{\mathcal{AC}(\lfloor \rho^\lambda \rfloor, \lfloor \sqrt{\lambda} \rfloor)} = \infty,$$

which implies that $(r_*^\lambda, q_*^\lambda)$ cannot be optimal, a contradiction. Thus, the proposition holds.

■

Now we show that $(r_*^\lambda, q_*^\lambda)$ satisfies Condition 2, which leads to (59) and (60).

Proposition 4 *The sequence of optimal $(r_*^\lambda, q_*^\lambda)$-policies for system-$\mathcal{S}^\lambda$ satisfies Condition* 2. *Hence* (59)-(60) *hold.*

Proof : According to the definition of Condition 2, it is sufficient to show

$$\Big|\lim_{\lambda\to\infty} z^\lambda(r_*^\lambda, q_*^\lambda)\Big| < \infty. \tag{78}$$

To that end, we first show that

$$\Big|\overline{\lim_{\lambda\to\infty}}\, z^\lambda(r_*^\lambda, q_*^\lambda)\Big| < \infty. \tag{79}$$

Suppose contrariwise that this does not hold. Then we have two possible cases:

$$\text{Case A:}\quad \underline{\lim}_{\lambda\to\infty}\, z^\lambda(r_*^\lambda, q_*^\lambda) = -\infty; \qquad \text{Case B:}\quad \overline{\lim_{\lambda\to\infty}}\, z^\lambda(r_*^\lambda, q_*^\lambda) = +\infty. \tag{80}$$

First, consider Case A. In view of Proposition 3, we have that if $K > 0$,

$$\underline{\lim}_{\lambda\to\infty}\, z^\lambda(i, q_*^\lambda) = -\infty \text{ for } i = r_*^\lambda + 1, \cdots, r_*^\lambda + q_*^\lambda. \tag{81}$$

If $K = 0$, by $q_*^\lambda = 1$, (81) also holds under Case A. Then there exists a subsequence $\{\lambda_k : k \geq 1\}$ such that $\lim_{k\to\infty} z^{\lambda_k}(i, q_*^\lambda) = -\infty$. We still write this subsequence as λ. By (36)-(37), for any policy (r^λ, q^λ),

$$\begin{aligned}
\mathsf{E}\left[\hat{G}\left(IN^\lambda\right)\right] &= \sum_{i=r^\lambda+1}^{r^\lambda+q^\lambda} \mathsf{E}\left[I\Big\{J^\lambda = r^\lambda + q^\lambda - i\Big\} \times \gamma^\lambda(q^\lambda) \times \hat{G}\left(z^\lambda(i, q^\lambda) - Y^\lambda(q^\lambda)\right)\right] \\
&= \gamma^\lambda(q^\lambda) \sum_{i=r^\lambda+1}^{r^\lambda+q^\lambda} \mathsf{E}\left[I\Big\{J^\lambda = r^\lambda + q^\lambda - i\Big\}\hat{G}\left(z^\lambda(i, q^\lambda) - Y^\lambda(q^\lambda)\right)\right]. \qquad (82)
\end{aligned}$$

We first consider each summand. Note that

$$\begin{aligned}
&\mathsf{E}\left[I\Big\{J^\lambda = r^\lambda + q^\lambda - i\Big\} \times \hat{G}\left(z^\lambda(i, q^\lambda) - Y^\lambda(q^\lambda)\right)\right] \\
&\quad = \int_{-\infty}^{\infty} \hat{G}\left(z^\lambda(i, q^\lambda) - y\right) \mathsf{dPr}\Big(Y^\lambda(q^\lambda) \leq y, J^\lambda = r^\lambda + q^\lambda - i\Big). \qquad (83)
\end{aligned}$$

By the first part of Theorem 1 and Theorem 1.A.3 (a) in Shaked and Shanthikumar (2007), we know that

$$\int_{-\infty}^{\infty} p\cdot\Big(z^\lambda(i,q^\lambda)-y\Big)^{-}\,\mathsf{dPr}\Big(Y^\lambda(q^\lambda)\le y, J^\lambda=r^\lambda+q^\lambda-i\Big)$$
$$\ge \frac{1}{q^\lambda}\int_{-\infty}^{\infty} p\cdot\Big(z^\lambda(i,q^\lambda)-y\Big)^{-}\,\mathsf{dPr}\Big(Y^\lambda(q^\lambda)-\beta^\lambda(q^\lambda)\le y\Big), \tag{84}$$
$$\int_{-\infty}^{\infty} h\cdot\Big(z^\lambda(i,q^\lambda)-y\Big)^{+}\,\mathsf{dPr}\Big(Y^\lambda(q^\lambda)\le y, J^\lambda=r^\lambda+q^\lambda-i\Big)$$
$$\ge \frac{1}{q^\lambda}\int_{-\infty}^{\infty} h\cdot\Big(z^\lambda(i,q^\lambda)-y\Big)^{+}\,\mathsf{dPr}\Big(Y^\lambda(q^\lambda)+\beta^\lambda(q^\lambda)\le y\Big). \tag{85}$$

Combining (84)-(85) yields

$$\int_{-\infty}^{\infty} \hat{G}\left(z^\lambda(i,q^\lambda)-y\right)\mathsf{dPr}\Big(Y^\lambda(q^\lambda)\le y, J^\lambda=r^\lambda+q^\lambda-i\Big) \tag{86}$$
$$\ge \frac{1}{q^\lambda}\mathsf{E}\Big[h\cdot\Big(z^\lambda(i,q^\lambda)-\beta^\lambda(q^\lambda)-Y^\lambda(q^\lambda)\Big)^{+}+p\cdot\Big(Y^\lambda(q^\lambda)-\beta^\lambda(q^\lambda)-z^\lambda(i,q^\lambda)\Big)^{+}\Big].$$

Considering policy $(r_*^\lambda, q_*^\lambda)$, we have, by Proposition 3, that for $i=r_*^\lambda+1,\cdots,r_*^\lambda+q_*^\lambda$,

$$\begin{aligned}
&\lim_{\lambda\to\infty}\mathsf{E}\Big[h\cdot\Big(z^\lambda(i,q_*^\lambda)-\beta^\lambda(q_*^\lambda)-Y^\lambda(q_*^\lambda)\Big)^{+}+p\cdot\Big(Y^\lambda(q_*^\lambda)-\beta^\lambda(q_*^\lambda)-z^\lambda(i,q_*^\lambda)\Big)^{+}\Big]\\
&\ge \lim_{\lambda\to\infty}\mathsf{E}\Big[p\cdot\Big(Y^\lambda(q_*^\lambda)-\beta^\lambda(q_*^\lambda)-z^\lambda(i,q_*^\lambda)\Big)^{+}\Big]\\
&\ge \lim_{\lambda\to\infty}\mathsf{E}\Big[p\cdot\Big(-\beta^\lambda(q_*^\lambda)-z^\lambda(i,q_*^\lambda)\Big)^{+}\times I\Big\{Y^\lambda(q_*^\lambda)\le 0\Big\}\Big]\\
&=\frac{p}{2}\times\lim_{\lambda\to\infty}\Big(-\beta^\lambda(q_*^\lambda)-z^\lambda(i,q_*^\lambda)\Big)^{+}\quad\text{(by Lemma 2)}\\
&=\infty.\quad\text{(by (81))}
\end{aligned}$$

It follows from (86) that for policy $(r_*^\lambda, q_*^\lambda)$,

$$\sum_{i=r_*^\lambda+1}^{r_*^\lambda+q_*^\lambda}\mathsf{E}\left[I\Big\{J^\lambda=r_+^\lambda+q_*^\lambda-i\Big\}\hat{G}\left(z^\lambda(i,q_*^\lambda)-Y^\lambda(q_*^\lambda)\right)\right]\to\infty\ \text{ as }\lambda\to\infty. \tag{87}$$

On the other hand, consider another policy $(r_0^\lambda, q_*^\lambda)$ with $r_0^\lambda=\Big\lfloor \rho^\lambda+\gamma^\lambda(q_*^\lambda)\Big\rfloor$. It is direct to verify that the sequence of $(r_0^\lambda, q_*^\lambda)$-policies satisfies Condition 2. Furthermore, by Proposition 3, the sequence of ordering quantities $\{q_*^\lambda\}$ satisfies Condition 1. Similar to the proof of (38) in Lemma 3, we can, by Conditions 1–2, show that for $i=r_0^\lambda+1,\cdots,r_0^\lambda+q_*^\lambda$,

$$\begin{aligned}
&\lim_{\lambda\to\infty}\mathsf{E}\Big[h\cdot\Big(z^\lambda(i,q_*^\lambda)+\beta^\lambda(q_*^\lambda)-Y^\lambda(q_*^\lambda)\Big)^{+}+p\cdot\Big(Y^\lambda(q_*^\lambda)+\beta^\lambda(q_*^\lambda)-z^\lambda(i,q_*^\lambda)\Big)^{+}\Big]\\
&=\lim_{\lambda\to\infty}\mathsf{E}\Big[h\cdot\Big(z^\lambda(i,q_*^\lambda)-\beta^\lambda(q_*^\lambda)-Y^\lambda(q_*^\lambda)\Big)^{+}+p\cdot\Big(Y^\lambda(q_*^\lambda)-\beta^\lambda(q_*^\lambda)-z^\lambda(i,q_*^\lambda)\Big)^{+}\Big]\\
&=\lim_{\lambda\to\infty}C\left(z^\lambda(i,q_*^\lambda)\right).
\end{aligned}\tag{88}$$

Combining (83)-(88) yields that for $i = r_0^\lambda + 1, \cdots, r_0^\lambda + q_*^\lambda$,

$$\lim_{\lambda\to\infty} q_*^\lambda \cdot \mathsf{E}\left[I\Big\{J^\lambda = r^\lambda + q_*^\lambda - i\Big\} \times \hat{G}\left(z^\lambda(i, q_*^\lambda) - Y^\lambda(q_*^\lambda)\right)\right] = \lim_{\lambda\to\infty} C\left(z^\lambda(i, q_*^\lambda)\right) < \infty. \quad (89)$$

Thus, from (87) and (89),

$$\frac{\sum_{i=r_*^\lambda+1}^{r_*^\lambda+q_*^\lambda} \mathsf{E}\left[I\Big\{J^\lambda = r_*^\lambda + q_*^\lambda - i\Big\}\hat{G}\left(z^\lambda(i, q_*^\lambda) - Y^\lambda(q_*^\lambda)\right)\right]}{\sum_{i=r_0^\lambda+1}^{r_0^\lambda+q_*^\lambda} \mathsf{E}\left[I\Big\{J^\lambda = r_0^\lambda + q_*^\lambda - i\Big\}\hat{G}\left(z^\lambda(i, q_*^\lambda) - Y^\lambda(q_*^\lambda)\right)\right]} \to \infty \quad \text{as } \lambda \to \infty.$$

which, by (9) and (36), contradicts the optimality of $(r_*^\lambda, q_*^\lambda)$. Therefore, Case A does not hold. Similarly, we can show Case B does not hold also. Hence (79) is proved.

To prove (78), with the help of (79), it is sufficient to show that for any convergent subsequence of $z^\lambda(r_*^\lambda, q_*^\lambda)$ (for the sake of notation simplicity, we still write it as $z^\lambda(r_*^\lambda, q_*^\lambda)$), its limit is always z_*. That is, we only need to prove

$$\lim_{\lambda\to\infty} z^\lambda(r_*^\lambda, q_*^\lambda) = z_*. \quad (90)$$

The convergence of the subsequence of $z^\lambda(r_*^\lambda, q_*^\lambda)$ implies that its corresponding subsequence of $(r_*^\lambda, q_*^\lambda)$ satisfies Condition 2. In view of Proposition 3, we know that $(r_*^\lambda, q_*^\lambda)$ satisfies Conditions 1–2 in Lemma 4. Thus, by Lemma 4,

$$\lim_{\lambda\to\infty} \frac{\mathcal{AC}(r_*^\lambda, q_*^\lambda)}{\widetilde{\mathcal{AC}}(r_*^\lambda, q_*^\lambda)} = 1. \quad (91)$$

Using Proposition 1, we know that the sequence of $(\tilde{r}_*^\lambda, \tilde{q}_*^\lambda)$-policies satisfies Conditions 1–2 with $\lim_{\lambda\to\infty} \tilde{q}_*^\lambda = \infty$. It follows from Lemma 4 that

$$\lim_{\lambda\to\infty} \frac{\mathcal{AC}(\tilde{r}_*^\lambda, \tilde{q}_*^\lambda)}{\widetilde{\mathcal{AC}}(\tilde{r}_*^\lambda, \tilde{q}_*^\lambda)} = 1. \quad (92)$$

On the other hand, by the optimality of $(r_*^\lambda, q_*^\lambda)$ for System-$\mathcal{S}^\lambda$ and the optimality $(\tilde{r}_*^\lambda, \tilde{q}_*^\lambda)$ for System-$\tilde{\mathcal{S}}^\lambda$, we have $\mathcal{AC}(r_*^\lambda, q_*^\lambda) \le \mathcal{AC}(\tilde{r}_*^\lambda, \tilde{q}_*^\lambda)$ and $\widetilde{\mathcal{AC}}(r_*^\lambda, q_*^\lambda) \ge \widetilde{\mathcal{AC}}(\tilde{r}_*^\lambda, \tilde{q}_*^\lambda)$. Hence from (91) and (92),

$$\lim_{\lambda\to\infty} \frac{\widetilde{\mathcal{AC}}(r_*^\lambda, q_*^\lambda)}{\widetilde{\mathcal{AC}}(\tilde{r}_*^\lambda, \tilde{q}_*^\lambda)} = 1. \quad (93)$$

With the help of Proposition 2, we, by (93), know that $\lim_{\lambda\to\infty} z^\lambda(r_*^\lambda, q_*^\lambda) = z_*$, which proves (90). This implies (78). The second part of the proposition ((59) and (60)) directly follows from (91) and (93). ■

Notice that in the proofs of Proposition 3, (78) (verify the sequence of $(r_*^\lambda, q_*^\lambda)$ satisfies Condition 2 in Proposition 4), Theorem 1 is not used for $(r_*^\lambda, q_*^\lambda)$ policy. Also the normal approximation for $\mathcal{AC}(r_*^\lambda, q_*^\lambda)$ is not used. With all the above preparations, we are now ready to show Part A of Theorem 2.

Proof : [**of Theorem 2 (Random Leadtimes)**] First, consider the case $K > 0$. Note that

$$\frac{\widetilde{\mathcal{AC}}(r_*^\lambda, q_*^\lambda)}{\widetilde{\mathcal{AC}}(\tilde{r}_*^\lambda, \tilde{q}_*^\lambda)} \geq \frac{\min_{r^\lambda} \widetilde{\mathcal{AC}}(r^\lambda, q_*^\lambda)}{\widetilde{\mathcal{AC}}(\tilde{r}_*^\lambda, \tilde{q}_*^\lambda)} \geq 1.$$

By (60),

$$\lim_{\lambda\to\infty} \frac{\min_{r^\lambda} \widetilde{\mathcal{AC}}(r^\lambda, q_*^\lambda)}{\widetilde{\mathcal{AC}}(\tilde{r}_*^\lambda, \tilde{q}_*^\lambda)} = 1.$$

This, by Proposition 2 (i), implies that $\lim_{\lambda\to\infty} q_*^\lambda/\tilde{q}_*^\lambda = 1$. Thus, by Proposition 1, we have (A.i); (A.ii) is directly given Proposition 2 (ii) and $\lim_{\lambda\to\infty} q_*^\lambda/\tilde{q}_*^\lambda = 1$; and (A.iii) immediately follows from (40) with $(r^\lambda, q^\lambda) = (\tilde{r}_*^\lambda, \tilde{q}_*^\lambda)$ and (92).

If $K = 0$, (A.i) follows from our assumption; (A.ii) and (A.iii) directly follow from (36)-(38), the definition of z_*, and Proposition 4. ■

5.3 Analysis: Constant Leadtimes

Similar to the random leadtime case, the proof of Part B of Theorem 2 also consists of three steps. Due to space constraint, we will only provide an outline of the analysis and leave the details to the Appendix.

Starting with (54), Step 1 establishes the asymptotic behavior for the optimal order quantity, reorder point and cost of System-$\tilde{\mathcal{S}}_c^\lambda$. This asymptotic behavior is the same as what we want to establish for System-$\mathcal{S}_c^\lambda$. Formally,

Proposition 5 *For System-$\tilde{\mathcal{S}}_c^\lambda$, the optimal policy $(\tilde{r}_{*c}^\lambda, \tilde{q}_{*c}^\lambda)$ and cost $\widetilde{\mathcal{AC}}_c(\tilde{r}_{*c}^\lambda, \tilde{q}_{*c}^\lambda)$ possess properties* (B.i)-(B.iii) *in Theorem* 2.

Step 2 proves the uniqueness for the first and second leading terms of the asymptotic optimal reorder point $(\tilde{r}_{*c}^\lambda)$, and the uniqueness for the leading term of the asymptotic optimal ordering quantity $(\tilde{q}_{*c}^\lambda)$ and cost of System-$\tilde{\mathcal{S}}_c^\lambda$. To characterize the uniqueness of $(\tilde{r}_{*c}^\lambda, \tilde{q}_{*c}^\lambda)$, we first define

$$\varpi^\lambda = \frac{\tilde{r}^\lambda - \rho^\lambda}{z_*\theta\sqrt{\rho^\lambda} - \alpha\tau\sqrt{\lambda}}.$$

Similar to Proposition 2, we have

Proposition 6 *Let* $\Delta_c(\lambda) = \tilde{q}^\lambda / \tilde{q}^\lambda_{*c}$.

(i) *If* $\overline{\lim}_{\lambda\to\infty} \Delta_c(\lambda) \neq 1$ *or* $\underline{\lim}_{\lambda\to\infty} \Delta_c(\lambda) \neq 1$*, then*

$$\overline{\lim_{\lambda\to\infty}} \frac{\min_{r^\lambda} \widetilde{\mathcal{AC}}_c(r^\lambda, \tilde{q}^\lambda)}{\widetilde{\mathcal{AC}}_c(\tilde{r}^\lambda_{*c}, \tilde{q}^\lambda_{*c})} > 1.$$

(ii) *Assume that* $\lim_{\lambda\to\infty} \Delta_c(\lambda) = 1$ *and* $\overline{\lim}_{\lambda\to\infty} |\varpi^\lambda| < \infty$*. If* $\overline{\lim}_{\lambda\to\infty} \varpi^\lambda \neq 1$ *or* $\underline{\lim}_{\lambda\to\infty} \varpi^\lambda \neq 1$*, then*

$$\overline{\lim_{\lambda\to\infty}} \frac{\widetilde{\mathcal{AC}}(\tilde{r}^\lambda, \tilde{q}^\lambda)}{\widetilde{\mathcal{AC}}(\tilde{r}^\lambda_{*c}, \tilde{q}^\lambda_{*c})} > 1.$$

Step 3 shows that the optimal policy of system-$\mathcal{S}_c^\lambda$ satisfies Conditions 1–3 which is needed in Lemma 5. Formally,

Proposition 7 *The sequence of optimal* $(r^\lambda_{*c}, q^\lambda_{*c})$*-policies for System-*$\mathcal{S}_c^\lambda$*,* (i) *if* $K > 0$*, then* $\underline{\lim}_{\lambda\to\infty} q^\lambda_{*c} = \infty$ *and* $\overline{\lim}_{\lambda\to\infty} q^\lambda_{*c}/\sqrt{\lambda} < \infty$; (ii) *there exists a constant* M *such that*

$$\Big|\lim_{\lambda\to\infty} z_c^\lambda(i)\Big| \le M, \ \ i = r^\lambda_{*c} + 1, \cdots, r^\lambda_{*c} + q^\lambda_{*c}.$$

With Propositions 5-7 in hand, we can prove Part B of Theorem 2.

Proof : [**of Theorem 2 for Constant Leadtimes**] Using Propositions 6-7, going along the line of the proof for random leadtime case, we can prove the constant leadtime case. Here the details are omitted. ■

6 (S, T) System with Constant Leadtimes

In this section we consider an (S, T) inventory system, where S is the order-up-to level and T the review period. In other words, we review the inventory position $IP(t)$ every T periods. If, upon review, $IP(t)$ is below S, then order enough to bring $IP(t)$ back to S; otherwise, do nothing. We assume full backlogging and a constant leadtime $1/\mu$. (We restrict to constant leadtimes here because this is the only case we know how to formulate the cost function.) Everything else (i.e., the demand process and cost structure) is the same as described in Section 2.

Let $\{A^\lambda(t) : t \ge 0\}$ be the renewal process generated by $\{\frac{\xi_n}{\lambda} : n \ge 1\}$. That is,

$$A^\lambda(t) \ = \ \max\left\{k \ : \ \frac{\xi_1}{\lambda} + \frac{\xi_2}{\lambda} + \cdots + \frac{\xi_k}{\lambda} \le t\right\}, \ \ t \ge 0.$$

Denote $A^\lambda_\infty(t) = \lim_{s\to\infty}[A^\lambda(s+t) - A^\lambda(s)]$. The objective is to minimize the long-run average total costs per unit time (see Rao 2003):

$$\min_{S,T} \overline{\mathcal{AC}}(S,T) := \frac{1}{T}\Big[K\cdot \mathsf{Pr}(A^\lambda_\infty(T) > 0) \\ + \mathsf{E}\int_{1/\mu}^{T+1/\mu}\Big(h\cdot(S - A^\lambda_\infty(t))^+ + p\cdot(S - A^\lambda_\infty(t))^-\Big)\,\mathsf{d}t\Big]. \quad (94)$$

Let $\{\overline{A}^\lambda(t) : t \geq 0\}$ be the delay renewal process generated by $\{\frac{\xi_n}{\lambda} : n \geq 1\}$. That is,

$$\overline{A}^\lambda(t) = \max\Big\{k \,:\, \frac{\bar{\xi}_1}{\lambda} + \frac{\xi_2}{\lambda} + \cdots + \frac{\xi_k}{\lambda} \leq t\Big\}, \quad t \geq 0. \quad (95)$$

$$\mathsf{Pr}(\bar{\xi}_1 \leq x) = \int_0^x (1 - \mathsf{Pr}(\xi_1 \leq t))\,\mathsf{d}t, \quad (96)$$

From the renewal theory (see Theorem 3.5.2 on p.131, Ross 1996), we have

$$\{A^\lambda_\infty(t) : t \geq 0\} \text{ and } \{\overline{A}^\lambda(t) : t \geq 0\} \text{ have the same distribution}$$

Hence, (94) can be rewritten as

$$\min_{S,T} \overline{\mathcal{AC}}(S,T) = \frac{1}{T}\Big[K\cdot \mathsf{Pr}(\overline{A}^\lambda(T) > 0) \\ + \mathsf{E}\int_{1/\mu}^{T+1/\mu}\Big(h\cdot(S - \overline{A}^\lambda(t))^+ + p\cdot(S - \overline{A}^\lambda(t))^-\Big)\,\mathsf{d}t\Big]. \quad (97)$$

Denote

$$Z^\lambda(t) = \frac{\overline{A}^\lambda(t) - \lambda t}{\theta\sqrt{\lambda t}}, \quad w^\lambda(t) = \frac{S - \lambda t}{\theta\sqrt{\lambda t}}. \quad (98)$$

Note that $w^\lambda(t)$ depends also on S. Sometimes we may write $w^\lambda(S,t)$ to highlight this dependence. By the invariance principle of the renewal process (see Theorem 14.6 on p.154, Billingsley 1999), we have that as $\lambda \to \infty$,

$$\sqrt{t}Z^\lambda(t) \text{ converges in distribution to a standard Brownian motion } \{\overline{A}(t) : t \geq 0\}. \quad (99)$$

By Theorem 9.1 on p.100 in Gut (2009), there exists a constant $\overline{M}$ such that for $t \in [\frac{1}{\mu}, \frac{1}{\mu} + T]$ and large enough λ,

$$\mathsf{E}\left(Z^\lambda(t)\right)^2 < \overline{M}.$$

This implies that for each fixed T and S,

$$\mathsf{E}\int_{1/\mu}^{T+1/\mu}\Big[h\cdot\left(w^\lambda(t) - Z^\lambda(t)\right)^+ + p\cdot\left(w^\lambda(t) - Z^\lambda(t)\right)^-\Big]\mathsf{d}t \to \int_{1/\mu}^{T+1/\mu} C\left(w^\lambda(t)\right)\mathsf{d}t.$$

Noticing (95)-(96), we have that $\lim_{\lambda\to\infty} \mathsf{Pr}(\overline{A}^{\lambda}(T) > 0) = 1$ for any $T > 0$. Hence it follows from (97)-(99) that, similar to the (r, q) system, we consider an auxiliary (S, T) system given by

$$\widehat{\mathcal{AC}}(S,T) = \frac{1}{T}\left(K + \int_{1/\mu}^{T+1/\mu} C\left(w^{\lambda}(t)\right)\cdot \theta\sqrt{\lambda t}\mathsf{d}t\right). \tag{100}$$

For fixed T, by the first-order condition, the optimal S to (100) is given by

$$\int_{1/\mu}^{T+1/\mu} \left[h\cdot\left(1-\Phi\left(-w^{\lambda}(t)\right)\right) - p\cdot\left(1-\Phi\left(w^{\lambda}(t)\right)\right)\right]\mathsf{d}t = 0. \tag{101}$$

Making integral variable transformation by $\lambda(t - 1/\mu) = x$, (101) can be written as

$$\frac{1}{\lambda T}\int_0^{\lambda T} \Phi\left(w^{\lambda}\left(\frac{1}{\mu}+\frac{x}{\lambda}\right)\right)\mathsf{d}x = \frac{p}{p+h}. \tag{102}$$

It is direct to verify that $\Phi\left(w^{\lambda}\left(\frac{1}{\mu}+\frac{x}{\lambda}\right)\right)$ is a decreasing function of x on the interval $[0,\infty)$. Thus, for $x \in [0, \lambda T]$,

$$\Phi\left(w^{\lambda}\left(\frac{1}{\mu}\right)\right) \geq \Phi\left(w^{\lambda}\left(\frac{1}{\mu}+\frac{x}{\lambda}\right)\right) \geq \Phi\left(w^{\lambda}\left(T+\frac{1}{\mu}\right)\right).$$

This, by (102), implies that

$$w^{\lambda}\left(\frac{1}{\mu}\right) > \Phi^{-1}\left(\frac{p}{p+h}\right) = z_* > w^{\lambda}\left(T+\frac{1}{\mu}\right).$$

For each given T, therefore, the optimal S (denoted by $\hat{S}_*^{\lambda}(T)$), that is, the solution to (101), can be written as

$$\hat{S}_*^{\lambda}(T) = \rho^{\lambda} + z_*\theta\sqrt{\rho^{\lambda}} + M^{\lambda}(T), \tag{103}$$

$$0 \leq M^{\lambda}(T) \leq \lambda T + z_*\theta\sqrt{\lambda T}. \tag{104}$$

Here the inequality $\sqrt{a+b} \leq \sqrt{a} + \sqrt{b}$ is applied for $a, b \geq 0$ in establishing (104). Plugging (103) into (100), by the first-order condition, we know that the optimal $\hat{T}_*^{\lambda}$ is given by

$$T\cdot C\left(w^{\lambda}\left(\hat{S}_*^{\lambda}(T), T+\frac{1}{\mu}\right)\right)\theta\sqrt{\lambda T+\rho^{\lambda}} - \left[K + \int_{1/\mu}^{T+1/\mu} C\left(w^{\lambda}(\hat{S}_*^{\lambda}(T),t)\right)\cdot\theta\sqrt{\lambda t}\mathsf{d}t\right] = 0 \tag{105}$$

This is equivalent to

$$T\cdot C\left(w^{\lambda}\left(\hat{S}_*^{\lambda}(T), T+\frac{1}{\mu}\right)\right)\theta\sqrt{T+\frac{1}{\mu}} - \int_{1/\mu}^{T+1/\mu} C\left(w^{\lambda}(\hat{S}_*^{\lambda}(T),t)\right)\cdot\theta\sqrt{t}\mathsf{d}t = \frac{K}{\sqrt{\lambda}}. \tag{106}$$

Thus, by (103) and the convexity of $C(\cdot)$, we can prove (the proof is provided in the Appendix) that

$$\sqrt{\lambda}\hat{T}_*^\lambda \quad \text{is bounded.} \tag{107}$$

In view of (104), we know that $M^\lambda(\hat{T}_*^\lambda)/\sqrt{\lambda}$ is also bounded. We pick up any two convergence sequences $\{\sqrt{\lambda_k}\hat{T}_*^{\lambda_k} \ : \ k \geq 1\}$ and $\{M^{\lambda_k}(\hat{T}_*^{\lambda_k})/\sqrt{\lambda_k} \ : \ k \geq 1\}$ from $\{\sqrt{\lambda}\hat{T}_*^\lambda \ : \ \lambda > 0\}$ and $\{M^\lambda(\hat{T}_*^\lambda)/\sqrt{\lambda} \ : \ \lambda > 0\}$ (again label them as λ). Let

$$\lim_{\lambda\to\infty} \sqrt{\lambda}\hat{T}_*^\lambda = \tau_1 \quad \text{and} \quad \lim_{\lambda\to\infty} \frac{M^\lambda(\hat{T}_*^\lambda)}{\sqrt{\lambda}} = \tau_2. \tag{108}$$

Then,

$$\begin{aligned}
&\lim_{\lambda\to\infty} \frac{1}{\lambda\hat{T}_*^\lambda} \int_0^{\lambda\hat{T}_*^\lambda} \Phi\left(w^\lambda\left(\hat{S}_*^\lambda(\hat{T}_*^\lambda), \frac{x}{\lambda} + \frac{1}{\mu}\right)\right) \mathsf{d}x \\
&= \lim_{\lambda\to\infty} \frac{1}{\sqrt{\lambda}\hat{T}_*^\lambda} \int_0^{\sqrt{\lambda}\hat{T}_*^\lambda} \Phi\left(w^\lambda\left(\hat{S}_*^\lambda(\hat{T}_*^\lambda), \frac{u}{\sqrt{\lambda}} + \frac{1}{\mu}\right)\right) \mathsf{d}u \quad (\text{setting } u = \tfrac{x}{\sqrt{\lambda}}) \\
&= \frac{1}{\tau_1} \int_0^{\tau_1} \Phi\left(z_* + \eta\tau_2 - \eta u\right) \mathsf{d}u.
\end{aligned} \tag{109}$$

By (103) we have

$$\lim_{\lambda\to\infty} \sqrt{\lambda}\hat{T}_*^\lambda \cdot C\left(w^\lambda(\hat{S}_*^\lambda(\hat{T}_*^\lambda), \hat{T}_*^\lambda + \frac{1}{\mu})\right) \theta\sqrt{\hat{T}_*^\lambda + \frac{1}{\mu}} = \frac{\tau_1}{\eta} C\left(z_* + \eta\tau_2 - \eta\tau_1\right), \tag{110}$$

$$\begin{aligned}
&\lim_{\lambda\to\infty} \sqrt{\lambda} \int_{1/\mu}^{\hat{T}_*^\lambda + 1/\mu} C\left(w^\lambda(\hat{S}_*^\lambda(\hat{T}_*^\lambda), t)\right) \cdot \theta\sqrt{t}\mathsf{d}t \\
&= \lim_{\lambda\to\infty} \sqrt{\lambda} \int_0^{\hat{T}_*^\lambda} C\left(w^\lambda\left(\hat{S}_*^\lambda(\hat{T}_*^\lambda), x + \frac{1}{\mu}\right)\right) \cdot \theta\sqrt{x + \frac{1}{\mu}}\mathsf{d}x \quad (\text{by setting } t - \tfrac{1}{\mu} = x) \\
&= \lim_{\lambda\to\infty} \int_0^{\sqrt{\lambda}\hat{T}_*^\lambda} C\left(\left(\frac{\hat{S}_*^\lambda(\hat{T}_*^\lambda)}{\sqrt{\lambda}} - \frac{\sqrt{\lambda}}{\mu} - u\right) \Big/ \theta\sqrt{\frac{u}{\sqrt{\lambda}} + \frac{1}{\mu}}\right) \cdot \theta\sqrt{\frac{u}{\sqrt{\lambda}} + \frac{1}{\mu}}\mathsf{d}u \\
&\qquad (\text{by setting } \sqrt{\lambda}x = u) \\
&= \eta^{-1} \int_0^{\tau_1} C\left(z_* + \eta\tau_2 - \eta u\right) \mathsf{d}u.
\end{aligned} \tag{111}$$

In view of (102) and (109), the limits given by (108) satisfy

$$\frac{1}{\tau_1} \int_0^{\tau_1} \Phi\left(z_* + \eta\tau_2 - \eta u\right) \mathsf{d}u = \frac{p}{p+h}. \tag{112}$$

Furthermore, in view of (106), (110)-(111) imply that the limits given by (108) also satisfy

$$\tau_1 \cdot C\left(z_* + \eta\tau_2 - \eta\tau_1\right) - \int_0^{\tau_1} C\left(z_* + \eta\tau_2 - \eta u\right) \mathsf{d}u = \eta K. \tag{113}$$

By the convexity of $C(\cdot)$, similar to Lemma 6, we can show that there exists a unique solution (τ_1, τ_2) to (112)-(113). Therefore we have

$$\lim_{\lambda\to\infty} \sqrt{\lambda}T_*^\lambda = \tau_1, \quad \lim_{\lambda\to\infty} \frac{M^\lambda(T_*^\lambda)}{\sqrt{\lambda}} = \tau_2, \tag{114}$$

and τ_1 and τ_2 are the solution to (112)-(113). Furthermore, from (105) and (110), we know that

$$\begin{aligned} &\frac{1}{\hat{T}_*^\lambda}\left(K + \int_{1/\mu}^{\hat{T}_*^\lambda + 1/\mu} C\left(w^\lambda\left(\hat{S}_*^\lambda(\hat{T}_*^\lambda), t\right)\right)\theta\sqrt{\lambda t}\mathsf{d}t\right) \\ &= C\left(w^\lambda\left(\hat{S}_*^\lambda(\hat{T}_*^\lambda), \hat{T}_*^\lambda + \frac{1}{\mu}\right)\right)\theta\sqrt{\lambda T_*^\lambda + \rho^\lambda} \\ &= \eta^{-1} \cdot C\left(z_* + \eta\tau_2 - \eta\tau_1\right)\sqrt{\lambda} + o(\sqrt{\lambda}). \end{aligned} \tag{115}$$

Summarizing (100), (103), and (114)-(115), we obtain

Proposition 8 *For the auxiliary (S, T) system given by* (100), *the optimal policy $(\hat{S}_*^\lambda, \hat{T}_*^\lambda)$ and cost $\widehat{\mathcal{AC}}(\hat{S}_*^\lambda, \hat{T}_*^\lambda)$ have the following relationships.*
(i) $\hat{S}_*^\lambda = \hat{S}_*^\lambda(\hat{T}_*^\lambda) = \lambda/\mu + z_*\theta\sqrt{\lambda/\mu} + \tau_2\sqrt{\lambda} + o(\sqrt{\lambda})$,
(ii) $\hat{T}_*^\lambda = \tau_1/\sqrt{\lambda} + o(1/\sqrt{\lambda})$,
(iii) $\widehat{\mathcal{AC}}(\hat{S}_*^\lambda, \hat{T}_*^\lambda) = \theta \cdot C\left(z_* + \eta(\tau_2 - \tau_1)\right)\sqrt{\lambda/\mu} + o(\sqrt{\lambda})$,
where τ_1 and τ_2 are the unique solutions to (112) *and* (113).

Similar to the proof of Theorem 2, by Proposition 8, we can establish the following results for the (S, T) system with constant leadtime.

Theorem 3 *For the constant leadtime, the the optimal policy $(S_*^\lambda, T_*^\lambda)$ and cost $\overline{\mathcal{AC}}(S_*^\lambda, T_*^\lambda)$ of (S, T)-system satisfy Proposition* 8 (i)-(iii) *respectively.*

Remark 5 By Theorems 2 and 3, we have

$$\lim_{\lambda\to\infty} \frac{\overline{\mathcal{AC}}(S_*^\lambda, T_*^\lambda) - \mathcal{AC}(r_*^\lambda, q_*^\lambda)}{\mathcal{AC}(r_*^\lambda, q_*^\lambda)} = \frac{C\left(z_* + \eta(\tau_2 - \tau_1)\right)}{C\left(z_* - \alpha c\eta\right)} - 1. \tag{116}$$

A similar bound is given by Rao (2003).

7 Conclusion

We have performed an asymptotic analysis of the (r, q) inventory system with a renewal demand process and i.i.d. stochastic leadtimes in heavy traffic. First, we have proved a previous conjecture that inventory position and inventory on-order are asymptotically independent. Second, we have established closed-form expressions for the asymptotically optimal policy parameters and system cost. These results reveal many interesting quantitative and qualitative effects of the system parameters on the optimal policy, such as demand and leadtime variability and fixed order cost. Most strikingly, we have shown that the well-known square-root relationship between the optimal order quantity and demand rate only holds for the special case of constant leadtimes. For the general i.i.d. random leadtimes, this relation is replaced by the cube root. Third, we have extended the analysis to periodic-review (S, T) systems with constant leadtimes. We hope our results and methods here can inspire future research to derive closed-form approximations of inventory policies for other inventory systems in order to sharpen intuition.

Acknowledgement: The authors would like to thank the Area Editor Costis Maglaras, the anonymous Associate Editor and referees for their valuable suggestions and corrections. This research was partially supported by the Chinese Natural Science Foundation through Grant No. 71390331 and the Singapore Management University through Grant No. C207MSS14B001 from the Ministry of Education Academic Research Fund Tier 1.

References

[1] Ang, M., Song, J.S., Wang, M., Zhang, H. 2013. On properties of discrete (r, q) and (s, T) inventory systems. *European Journal of Operational Research*, **229**, 95-105.

[2] Axsäter, S. 1996. Using the deterministic EOQ formula in stochastic inventory control. *Management Science*, **42**, 830-834.

[3] Benjaafar, S., Chen, D., ElHafsi, M. 2014. Optimal control of an inventory system with stochastic and independent leadtimes. *Working Paper*, Department of Industrial and Systems Engineering, University of Minnesota, MN 55455, USA.

[4] Billingsley, P. 1999. *Convergence of Probability Measures.* 2nd Edition, Wiley, New York.

[5] Borovkov, A.A. 1967. On limit laws for service processes in multi-channel systems. *Siberian Math. J.*, **8**, 746-763.

[6] Bradley, J., Robinson, L. 2005. Improved base-stock approximations for independent stochastic lead times with order crossover. *Manufacturing & Service Operations Management*, **7**, 319-329.

[7] Bradley, J., Robinson, L. 2008. Further improvements on base-stock approximations for independent stochastic lead times with order crossover. *Manufacturing & Service Operations Management*, **9**, 325-327.

[8] Chow, Y.S., Teicher, H. 2003. *Probability Theory.* 3rd Edition, Springer.

[9] Federgruen, A., Wang, M. 2013. Monotonicity properties of a class of stochastic inventory systems. *Annals of Operations Research*, **208**, 155-186.

[10] Federgruen, A., Zheng, Y.S. 1992. An efficient algorithm for computing an optimal (r, Q) policy in continuous review stochastic inventory systems. *Operations Research*, **40**(4), 808-813.

[11] Gallego, G. 1998. New bounds and heuristics for (q, R) policies. *Management Science*, **44**, 219-233.

[12] Galliher, H., Morse, P., Simond, M. 1959. Dynamics of two classes of continuous-review inventory systems. *Operations Research*, **7**, 362-384.

[13] Gut, A. 2009. *Stopped Random Walks.* 2nd Edition, Springer.

[14] Kulkarni, V., Yan, K. 2012. Production-inventory systems in stochastic environment and stochastic lead times. *Queueing Systems*, **70**, 207-231.

[15] Muharremoglu, A., Yang, N. 2010. Inventory management with an exogenous supply process. *Operations Research*, **58**, 111-129.

[16] Muthuraman, K., Seshadri, S., Wu, Q. 2015. Inventory management with Stochastic Lead Times. *Mathematics of Operations Research*, **40**, 302-327.

[17] Platt, D., Robinson, L., Freund, R. 1997. Tractable (Q, R) heuristic models for constrained service levels. *Management Science* **43**, 951-965.

[18] Rao, U.S. 2003. Properties of the periodic review (R, T) inventory control policy for stationary, stochastic demand. *Manufacturing and Service Operations Managment*, **5**, 37-53.

[19] Rudin, W. 1976. *Principles of Mathematical Analysis.* 3rd Edition, McGraw-Hill International Editions.

[20] Ross, S.M. 1996. *Stochastic Processes.* 2nd Edition, Wiley, New York.

[21] Sahin, I. 1983. On the continuous review (s, S) inventory model under compound renewal demand and random lead times. *Journal of Applied Probability*, **20**, 213-219.

[22] Scarf, H. 1958. Stationary operating characteristics of an inventory model with time lag. In K. Arrow, S. Karlin and H. Scarf (Eds.), *Studies in the Mathematica Theory of Inventory and Production*, Stanford University, Stanford, CA.

[23] Shaked, M., Shanthikumar, J.G. 2007. *Stochastic Orders*. Springer.

[24] Sigman, K. 1996. *Stationary Marked Point Processes: An Intuitive Approach*. Chapman and Hall, New York, London.

[25] Sigman, K., Whitt, W. 2011. Heavy-traffic limits for nearly deterministic queues: stationary distributions. *Queueing Systems*, **69**, 145-173.

[26] Simon, R.M. 1968. The uniform distribution of inventory position for continuous review (s, Q) policies. *RAND Corporation Report*, No. P-3938, Santa Monica, CA.

[27] Song, J.S. 2000. A note on assemble-to-order systems with batch ordering. *Management Science*, **46**, 739-743.

[28] Song, J.S., Zhang, H., Hou, Y., Wang, M. 2010. The effect of lead time and demand uncertainties in (r, q) inventory systems. *Operations Research*, **58**, 68-80.

[29] Song, J.S., Zipkin, P. 1996. The joint effect of leadtime variance and lot size in a parallel processing environmment. *Management Science*, **42**, 1352-1363.

[30] Tak*á*cs, L. 1958. On a coincidence problem in telephone traffic. *Acta Math. Acad. Sci. Hung.*, **9**, 45-80.

[31] Whitt W. 1992. Understanding the efficiency of multi-server service systems. *Management Science*, **38**, 708-723.

[32] Yamazaki, G., Sigman, K., Miyazawa. M. 1992. Moments in infinite channel queues. *Computers Math. Applic.*, **24**, 1-6.

[33] Zalkind, D. 1978. Order-level inventory systems with independent stochastic lead times. *Management Science*, **24**, 1384-1392.

[34] Zheng, Y.S. 1992. On properties of stochastic inventory systems. *Management Science*, **38**, 87-103.

[35] Zipkin, P. 1986. Stochastic leadtimes in continuous-time inventory models. *Naval Research Logistics Quarterly*, **33**, 763-774.

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

Appendix A:

Proof : **[of (9) on the expected long-run average cost]** Eq (9) is a well-known result under exogenous sequential leadtimes (see the proof of Theorem 6.5.1 on p.219 in Zipkin, 2000). Under the parallel supply systems considered here, we were unsuccessful in finding the original source for this expression. Therefore, we provided a rigorous proof here; an earlier proof may be hidden in some literature. First, from the elementary renewal theory (see Theorem 3.3.4 on p.107, Ross 1996), we have

$$\lim_{T\to\infty} \mathsf{E}A_q(T)/T = \lambda/q.$$

Thus to prove (9), it suffices to show that

$$\lim_{T\to\infty} \frac{1}{T}\mathsf{E}\int_0^T \hat{G}(IN(t))\mathsf{d}t = \mathsf{E}[\hat{G}(IN)]. \tag{A-1}$$

Noticing that for each t, $\hat{G}(IN(t))$ is nonnegative, from Theorem 2 on p.186, Chow and Teicher (2003), we have

$$\mathsf{E}\int_0^T \hat{G}(IN(t))\ \mathsf{d}t = \int_0^T \mathsf{E}[\hat{G}(IN(t))]\ \mathsf{d}t.$$

Therefore, to prove (A-1), it is sufficient to show that

$$\lim_{t\to\infty} \mathsf{E}[\hat{G}(IN(t))] = \mathsf{E}[\hat{G}(IN)]. \tag{A-2}$$

In view of (7)-(8), we first consider the joint distribution of $J(t)$ and $N(t)$. Similar to the proof of Theorem 1, with the help of the time-stationary point process framework discussed in Sigman (1996), we consider the double infinite sequence $\{t_k : k = \pm 1, \pm 2, \cdots\}$ with $t_{-1} \le 0 < t_1$. The inventory system has been in operation since the infinite past and demand arrivals follow the sequence $\{t_k : k = \pm 1, \pm 2, \cdots\}$. Denote the leadtime experienced by the nth order by $\{\zeta_n : n = \pm 1, \pm 2, \cdots\}$. From the time-stationary point process framework, we have

$$(J(t), N(t)) \ \text{ converges to } \ (J, N) \ \text{ in distribution,}$$

and (J, N) and $(J(0), N(0))$ have the same distribution. Using (7)-(8), we have

$$\hat{G}(IN(t)) \le (p+h)\cdot\Big(2(r+q) + qN(t)\Big). \tag{A-3}$$

Note that $N(t)$ is the queue length of $GI/GI/\infty$ at time t with the interarrival times $\{t_{nq} - t_{(n-1)q} : n \ge 1\}$ $(t_0 = 0)$ and service time distribution given by $F(\cdot)$. By the proof of Theorem 2 on p.50 in Takács (1958), we know that there exists a constant M such that for any t, $\mathsf{E}N(t) \le M$. This combining with (A-3) gives that

$$\mathsf{E}[\hat{G}(IN(t))] \le (p+h)\cdot\Big(2(r+q) + qM\Big).$$

This, by Theorem 2 on p.276 in Chow and Teicher (2003), implies that (A-2) holds. ■

Proof : **[of Lemma 3]** First we prove (38). By Proposition 3.1 in Yamazaki et al. (1992),

$$\mathsf{E}\Big(N^\lambda - \frac{\rho^\lambda}{q^\lambda}\Big)^2 \le \frac{\rho^\lambda}{q^\lambda} + 2\frac{\lambda}{q^\lambda}\cdot\frac{\theta}{\sqrt{q^\lambda}}\mathsf{E}L_{(2)}.$$

By the definition of $Y^\lambda(q^\lambda)$ given by (16) and $\left|\lim_{\lambda\to\infty} z^\lambda(i^\lambda, q^\lambda)\right| < \infty$, there exists a constant M such that

$$\mathsf{E}\left(z^\lambda(i^\lambda, q^\lambda) - Y^\lambda(q^\lambda)\right)^2 \le M.$$

Thus, the function $\hat{G}(\cdot)$ is uniformly integrable relative to the sequence of distribution functions given by $\left(z^\lambda(i^\lambda, q^\lambda) - Y^\lambda(q^\lambda)\right)$. Hence, (38) directly follows from Theorem 2 on p.276 in Chow and Teicher (2003) and Lemma 2 (as Condition 1 holds and $\lim_{\lambda\to\infty} z^\lambda(i^\lambda, q^\lambda)$ exists).

Now we prove (39). This approximation result is mentioned by Zheng (1992) (see p.89, Zheng 1992). It may be hidden in some textbooks. For the completeness, here we give a proof. It suffices to show that for any $\varepsilon > 0$, there exists an Λ such that for $\lambda > \Lambda$,

$$\Big|\sum_{i=r^\lambda+1}^{r^\lambda+q^\lambda} C\left(z^\lambda(i, q^\lambda)\right) - \int_{r^\lambda}^{r^\lambda+q^\lambda} C\left(z^\lambda(x, q^\lambda)\right)\mathsf{d}x\Big| \le \varepsilon \times \int_{r^\lambda}^{r^\lambda+q^\lambda} C\left(z^\lambda(x, q^\lambda)\right)\mathsf{d}x. \tag{A-4}$$

Note that

$$\begin{aligned}
&\sum_{i=r^\lambda+1}^{r^\lambda+q^\lambda} C\left(z^\lambda(i, q^\lambda)\right) - \int_{r^\lambda}^{r^\lambda+q^\lambda} C\left(z^\lambda(x, q^\lambda)\right)\mathsf{d}x \\
&\quad = \sum_{i=r^\lambda+1}^{r^\lambda+q^\lambda}\left[C\left(z^\lambda(i, q^\lambda)\right) - \int_{i-1}^{i} C\left(z^\lambda(x, q^\lambda)\right)\mathsf{d}x\right] \\
&\quad = \sum_{i=r^\lambda+1}^{r^\lambda+q^\lambda}\int_{i-1}^{i}\left[C\left(z^\lambda(i, q^\lambda)\right) - C\left(z^\lambda(x, q^\lambda)\right)\right]\mathsf{d}x.
\end{aligned} \tag{A-5}$$

Using (32) and the convexity of $C(\cdot)$, for $i = r^\lambda + 1, \cdots, r^\lambda + q^\lambda$ and $x \in [i-1,\ i]$, if $z_* \notin (z^\lambda(i, q^\lambda),\ z^\lambda(i+1, q^\lambda))$, then

$$\begin{aligned}
\left|C\left(z^\lambda(i, q^\lambda)\right) - C\left(z^\lambda(x, q^\lambda)\right)\right| &\le \left|C\left(z^\lambda(i+1, q^\lambda)\right) - C\left(z^\lambda(i, q^\lambda)\right)\right| \\
&= h \times \left|\Phi^1\left(-z^\lambda(i+1, q^\lambda)\right) - \Phi^1\left(-z^\lambda(i, q^\lambda)\right)\right| \\
&\quad + p \times \left|\Phi^1\left(z^\lambda(i+1, q^\lambda)\right) - \Phi^1\left(z^\lambda(i, q^\lambda)\right)\right| \\
&\le \frac{p}{\sigma^\lambda(q^\lambda)\sqrt{\lambda q^\lambda}} + \frac{h}{\sigma^\lambda(q^\lambda)\sqrt{\lambda q^\lambda}},
\end{aligned} \tag{A-6}$$

and if $z_* \in (z^\lambda(i, q^\lambda),\ z^\lambda(i+1, q^\lambda))$, then

$$\begin{aligned}
&\left|C\left(z^\lambda(i, q^\lambda)\right) - C\left(z^\lambda(x, q^\lambda)\right)\right| \\
&\quad \le \max\Big\{C\left(z^\lambda(i+1, q^\lambda)\right) - C(z_*), C\left(z^\lambda(i, q^\lambda)\right) - C(z_*)\Big\}.
\end{aligned} \tag{A-7}$$

By Conditions 1–2, there exists an Λ_0 such that for $\lambda > \Lambda_0$,

$$\left|z^\lambda(i, q^\lambda)\right| \leq M + 1, \; i = r^\lambda + 1, \cdots, r^\lambda + q^\lambda. \tag{A-8}$$

Thus, for $i = r^\lambda + 1, \cdots, r^\lambda + q^\lambda$,

$$\int_{i-1}^{i} C\left(z^\lambda(x, q^\lambda)\right) \mathsf{d}x \geq h \times \Phi^1(M+1) + p \times \Phi^1(M+1). \tag{A-9}$$

From the definition of $\sigma^\lambda(q^\lambda)$ given by (14) and (A-9), for any $\varepsilon > 0$, there exists an Λ_1 such that for $\lambda > \Lambda_1$,

$$\frac{p}{\sigma^\lambda(q^\lambda)\sqrt{\lambda q^\lambda}} + \frac{h}{\sigma^\lambda(q^\lambda)\sqrt{\lambda q^\lambda}} \leq \varepsilon \times \int_{i-1}^{i} C\left(z^\lambda(x, q^\lambda)\right) \mathsf{d}x. \tag{A-10}$$

Combining (A-6) and (A-10) yields that for i with $z_* \notin (z^\lambda(i, q^\lambda), \; z^\lambda(i+1, q^\lambda))$,

$$\int_{i-1}^{i} \left|C\left(z^\lambda(i, q^\lambda)\right) - C\left(z^\lambda(x, q^\lambda)\right)\right| \mathsf{d}x \leq \varepsilon \times \int_{i-1}^{i} C\left(z^\lambda(x, q^\lambda)\right) \mathsf{d}x. \tag{A-11}$$

With the help of (32), for i with $z_* \in (z^\lambda(i, q^\lambda), \; z^\lambda(i+1, q^\lambda))$, similarly, we can show that there exists an Λ_2 such that for $\lambda > \Lambda_2$,

$$\max\left\{C\left(z^\lambda(i+1, q^\lambda)\right) - C(z_*), C\left(z^\lambda(i, q^\lambda)\right) - C(z_*)\right\} \leq \varepsilon \times \int_{i-1}^{i} C\left(z^\lambda(x, q^\lambda)\right) \mathsf{d}x. \tag{A-12}$$

Combining (A-7) and (A-12) yields that for i with $z_* \in (z^\lambda(i, q^\lambda), \; z^\lambda(i+1, q^\lambda))$,

$$\int_{i-1}^{i} \left|C\left(z^\lambda(i, q^\lambda)\right) - C\left(z^\lambda(x, q^\lambda)\right)\right| \mathsf{d}x \leq \varepsilon \times \int_{i-1}^{i} C\left(z^\lambda(x, q^\lambda)\right) \mathsf{d}x. \tag{A-13}$$

Therefore, (A-4) holds for $\lambda > \max\{\Lambda_0, \Lambda_1, \Lambda_2\}$ directly from (A-5), (A-11) and (A-13). Thus the validity of the approximation given by (39) is proved. ■

Proof : **[of Lemma 4]** In view of (9) and (35), it suffices to show that

$$\lim_{\lambda\to\infty} \left(q^\lambda \cdot \mathsf{E}\left[\hat{G}\left(IN^\lambda\right)\right]\right) \Big/ \left(\gamma^\lambda(q^\lambda) \int_{r^\lambda}^{r^\lambda+q^\lambda} C\left(z^\lambda(x, q^\lambda)\right) \mathsf{d}x\right) = 1.$$

It follows from Lemma 3 that this is equivalent to show that

$$\lim_{\lambda\to\infty} \left(q^\lambda \cdot \mathsf{E}\left[\hat{G}\left(IN^\lambda\right)\right]\right) \Big/ \left(\gamma^\lambda(q^\lambda) \sum_{i=r^\lambda+1}^{r^\lambda+q^\lambda} C\left(z^\lambda(i, q^\lambda)\right)\right) = 1. \tag{A-14}$$

To prove (A-14), in view of (82), we only need to show that

$$\lim_{\lambda\to\infty} \left(q^\lambda \sum_{i=r^\lambda+1}^{r^\lambda+q^\lambda} \mathsf{E}\left[I\left\{J^\lambda = r^\lambda + q^\lambda - i\right\} \times \hat{G}\left(z^\lambda(i, q^\lambda) - Y^\lambda(q^\lambda)\right)\right]\right) \Big/ \left(\sum_{i=r^\lambda+1}^{r^\lambda+q^\lambda} C\left(z^\lambda(i, q^\lambda)\right)\right) = 1. \tag{A-15}$$

To that end, we first consider each summand. Similar to (83), we have that

$$\begin{aligned}
&\mathsf{E}\left[I\Big\{J^\lambda = r^\lambda + q^\lambda - i\Big\} \times \hat{G}\left(z^\lambda(i,q^\lambda) - Y^\lambda(q^\lambda)\right)\right] \\
&\quad = \int_{-\infty}^{\infty} \hat{G}\left(z^\lambda(i,q^\lambda) - y\right) \mathsf{d}\mathsf{Pr}\Big(Y^\lambda(q^\lambda) \le y, J^\lambda = r^\lambda + q^\lambda - i\Big). \qquad \text{(A-16)}
\end{aligned}$$

By the first part of Theorem 1 and Theorem 1.A.3 on p.6, Shaked and Shanthikumar (2007), we know that

$$\begin{aligned}
&\int_{-\infty}^{\infty} \hat{G}\left(z^\lambda(i,q^\lambda) - y\right) \mathsf{d}\mathsf{Pr}\Big(Y^\lambda(q^\lambda) \le y, J^\lambda = r^\lambda + q^\lambda - i\Big) \qquad \text{(A-17)} \\
&\quad \le \frac{1}{q^\lambda}\mathsf{E}\Big[h \cdot \Big(z^\lambda(i,q^\lambda) + \beta^\lambda(q^\lambda) - Y^\lambda(q^\lambda)\Big)^+ + p \cdot \Big(Y^\lambda(q^\lambda) + \beta^\lambda(q^\lambda) - z^\lambda(i,q^\lambda)\Big)^+\Big]; \\
&\int_{-\infty}^{\infty} \hat{G}\left(z^\lambda(i,q^\lambda) - y\right) \mathsf{d}\mathsf{Pr}\Big(Y^\lambda(q^\lambda) \le y, J^\lambda = r^\lambda + q^\lambda - i\Big) \qquad \text{(A-18)} \\
&\quad \ge \frac{1}{q^\lambda}\mathsf{E}\Big[h \cdot \Big(z^\lambda(i,q^\lambda) - \beta^\lambda(q^\lambda) - Y^\lambda(q^\lambda)\Big)^+ + p \cdot \Big(Y^\lambda(q^\lambda) - \beta^\lambda(q^\lambda) - z^\lambda(i,q^\lambda)\Big)^+\Big].
\end{aligned}$$

Similar to the proof of (38) in Lemma 3, we can, by Conditions 1–2, show that

$$\begin{aligned}
&\lim_{\lambda\to\infty} \mathsf{E}\Big[h \cdot \Big(z^\lambda(i,q^\lambda) + \beta^\lambda(q^\lambda) - Y^\lambda(q^\lambda)\Big)^+ + p \cdot \Big(Y^\lambda(q^\lambda) + \beta^\lambda(q^\lambda) - z^\lambda(i,q^\lambda)\Big)^+\Big] \\
&\quad = \lim_{\lambda\to\infty} \mathsf{E}\Big[h \cdot \Big(z^\lambda(i,q^\lambda) - \beta^\lambda(q^\lambda) - Y^\lambda(q^\lambda)\Big)^+ + p \cdot \Big(Y^\lambda(q^\lambda) - \beta^\lambda(q^\lambda) - z^\lambda(i,q^\lambda)\Big)^+\Big] \\
&\quad = \lim_{\lambda\to\infty} C\left(z^\lambda(i,q^\lambda)\right). \qquad \text{(A-19)}
\end{aligned}$$

Combining (A-16)-(A-19) yields that

$$\lim_{\lambda\to\infty} q^\lambda \cdot \mathsf{E}\left[I\Big\{J^\lambda = r^\lambda + q^\lambda - i\Big\} \times \hat{G}\left(z^\lambda(i,q^\lambda) - Y^\lambda(q^\lambda)\right)\right] = \lim_{\lambda\to\infty} C\left(z^\lambda(i,q^\lambda)\right),$$

which implies that (A-15) holds. Therefore, the lemma is proved. ■

Proof : **[of Lemma 6]** If $K > 0$, from (55), we know that $\tau \ne 0$. Thus for positive τ and η, by the strict convexity of $C(\cdot)$, we know that there exists a unique $\alpha \in (0,1)$ (write as $g(\tau)$) such that

$$C\left(z_* - g(\tau)\tau\eta\right) = C\left(z_* + (1 - g(\tau))\tau\eta\right).$$

Furthermore, this, by the strict convexity of $C(\cdot)$, implies that

$$\frac{\mathsf{d}\left(g(\tau)\tau\right)}{\mathsf{d}\tau} = \frac{C'(z_* + (1 - g(\tau))\tau\eta)}{C'(z_* + (1 - g(\tau))\tau\eta) - C'(z_* - g(\tau)\tau\eta)} \ne 0.$$

Plugging $g(\tau)$ into (55), we have

$$-\eta K + \tau \times C\left(z_* - g(\tau)\tau\eta\right) - \frac{1}{\eta}\int_{z_*-g(\tau)\tau\eta}^{z_*+(1-g(\tau))\tau\eta} C(y)\mathsf{d}y = 0.$$

Taking derivative on the left-hand side with respect to τ, we have

$$-\tau C'\left(z_*-g(\tau)\tau\eta\right)\cdot\frac{\mathsf{d}\left(g(\tau)\tau\right)}{\mathsf{d}\tau}\neq 0.$$

The existence of τ directly follows from the implicit function theorem (see Theorem 9.28 on p.224 of Rudin 1976).

Finally we show that $\tau\in(0,\infty)$. $\tau\neq 0$ directly follows from $K>0$, $\eta>0$, and (55). Suppose contrariwise that $\tau<0$. From (55), we have

$$\eta^2K=\tau\eta\times C\left(z_*-\alpha\tau\eta\right)-\int_{z_*-\alpha\tau\eta}^{z_*+(1-\alpha)\tau\eta}C(y)\mathsf{d}y.$$

This is equivalent to

$$\eta^2K=\int_{z_*-\alpha\tau\eta}^{z_*+(1-\alpha)\tau\eta}\Big(C\left(z_*-\alpha\tau\eta\right)-C(y)\Big)\mathsf{d}y. \tag{A-20}$$

By (56) and the convexity of $C(\cdot)$, and noticing that $z_*+(1-\alpha)\eta\tau<z_*-\alpha\eta\tau$ if $\tau<0$, we have that

$$C\left(z_*-\alpha\eta\tau\right)-C(y)\geq 0\quad\text{for } y\in[z_*+(1-\alpha)\eta\tau,z_*-\alpha\eta\tau].$$

This implies that

$$\int_{z_*-\alpha\eta\tau}^{z_*+(1-\alpha)\eta\tau}\Big(C\left(z_*-\alpha\eta\tau\right)-C(y)\Big)\mathsf{d}y\leq 0.$$

Thus we get a contradiction from (A-20) as $\eta^2K>0$. Hence, $\tau\in(0,\infty)$. Thus, the proof of the lemma is completed. ■

Proof : **[of Lemma 7]** By the definitions of $\sigma^\lambda(q^\lambda)$ and $\beta^\lambda(q^\lambda)$ given by (14)-(15), in view of (43), it is sufficient to consider the differentiability of $\kappa(q^\lambda)$. By the strictly convexity of $C(\cdot)$ (see (33)), we know the continuity of $\kappa(\cdot)$. Using (42), for any $\delta>0$,

$$\begin{aligned}&C\Big(z_*-\kappa(q^\lambda+\delta)\Big)-C\Big(z_*-\kappa(q^\lambda)\Big)\\&\quad=C\Big(z_*+\beta^\lambda(q^\lambda+\delta)-\kappa(q^\lambda+\delta)\Big)-C\Big(z_*+\beta^\lambda(q^\lambda)-\kappa(q^\lambda)\Big).\end{aligned}$$

This implies that

$$\begin{aligned}&\frac{\kappa(q^\lambda+\delta)-\kappa(q^\lambda)}{\delta}\times\Big[\frac{C\Big(z_*+\beta^\lambda(q^\lambda+\delta)-\kappa(q^\lambda+\delta)\Big)-C\Big(z_*+\beta^\lambda(q^\lambda)-\kappa(q^\lambda)\Big)}{\Big(\beta^\lambda(q^\lambda+\delta)-\kappa(q^\lambda+\delta)\Big)-\Big(\beta^\lambda(q^\lambda)-\kappa(q^\lambda)\Big)}\\&\qquad-\frac{C\Big(z_*-\kappa(q^\lambda+\delta)\Big)-C\Big(z_*-\kappa(q^\lambda)\Big)}{\kappa(q^\lambda)-\kappa(q^\lambda+\delta)}\Big]\\&=\frac{C\Big(z_*+\beta^\lambda(q^\lambda+\delta)-\kappa(q^\lambda+\delta)\Big)-C\Big(z_*+\beta^\lambda(q^\lambda)-\kappa(q^\lambda)\Big)}{\Big(\beta^\lambda(q^\lambda+\delta)-\kappa(q^\lambda+\delta)\Big)-\Big(\beta^\lambda(q^\lambda)-\kappa(q^\lambda)\Big)}\times\frac{\beta^\lambda(q^\lambda+\delta)-\beta^\lambda(q^\lambda)}{\delta}.\end{aligned} \tag{A-21}$$

Letting δ go to zero, by the continuity of $\beta^\lambda(\cdot)$ and $\kappa(\cdot)$, we know the right-hand side of (A-21) does converge to

$$C'\left(z_* + \beta^\lambda(q^\lambda) - \kappa(q^\lambda)\right) \times \frac{\mathsf{d}\beta^\lambda(q^\lambda)}{\mathsf{d}q^\lambda}.$$

Similarly, the second factor of the left-hand side of (A-21) does converge to

$$C'\left(z_* + \beta^\lambda(q^\lambda) - \kappa(q^\lambda)\right) - C'\left(z_* - \kappa(q^\lambda)\right).$$

By (43), the strictly convexity of $C(\cdot)$ and the definition of z_* given by (33), we have that for $q^\lambda > 0$,

$$C'\left(z_* + \beta^\lambda(q^\lambda) - \kappa(q^\lambda)\right) > 0 \text{ and } C'\left(z_* - \kappa(q^\lambda)\right) < 0.$$

Thus we know that the limit of the second factor of the left-hand side of (A-21) is positive. Hence we know the limit of the first factor of the left-hand side of (A-21) does exist, which gives the differentiability of $\kappa(\cdot)$. ■

Proof : [**of Lemma 8**] According to the definitions of $\kappa(\tilde{q}_*^\lambda)$ and $\overline{\kappa}(\tilde{q}_*^\lambda)$ given by (43), it suffices to show the first part of the lemma, namely,

$$\lim_{\lambda\to\infty} \tilde{q}_*^\lambda = \infty \quad \text{and} \quad \lim_{\lambda\to\infty} \frac{\tilde{q}_*^\lambda}{\lambda} = 0. \tag{A-22}$$

If the first equation does not hold, then, by (43), the right-hand side of (62) will go to zero while the left-hand side is fixed at $K > 0$. And if the second equation does not hold, the right-hand side of (62) will go to infinite while the left-hand side is fixed at K. And thus $\tilde{q}_*^\lambda$ cannot be a solution of (62) when (A-22) does not hold. Thus we have (A-22). This in turn implies the lemma. ■

Proof : [**of Proposition 5**] First we consider $K > 0$ case. Similar to the proof of Proposition 1, we need to establish the result similar to Lemma 8. Namely,

$$\lim_{\lambda\to\infty} \tilde{q}_{*c}^\lambda = \infty \quad \text{and} \quad \frac{\tilde{q}_{*c}^\lambda}{\sqrt{\lambda}} \quad \text{is bounded.} \tag{A-23}$$

If the first equation is not true, then there exists a subsequence $\{\lambda_k : k \geq 1\}$ such that

$$\lim_{k\to\infty} \lambda_k = \infty \quad \text{and} \quad \lim_{k\to\infty} \tilde{q}_{*c}^{\lambda_k} = a < \infty. \tag{A-24}$$

We still label the subsequence of (A-24) by λ. Under (A-24), by the definition of γ_c^λ given by (45), we have

$$\kappa_c(\tilde{q}_{*c}^\lambda) \to 0 \quad \text{and} \quad \overline{\kappa}_c(\tilde{q}_{*c}^\lambda) \to 0, \tag{A-25}$$

which implies

$$\lim_{\lambda\to\infty} C(z_* - \kappa_c(\tilde{q}_{*c}^\lambda)) = C_*. \tag{A-26}$$

This plus the mean-value theorem for integration yields

$$\frac{\gamma_c^\lambda}{\tilde{q}_{*c}^\lambda} \int_{z_*-\kappa_c(\tilde{q}_{*c}^\lambda)}^{z_*+\overline{\kappa}_c(\tilde{q}_{*c}^\lambda)} C(x)\mathsf{d}x \to C_*. \tag{A-27}$$

Combining (A-26)–(A-27) yields that the right-hand-side of (54) converges to zero. However, (A-24) implies that $\frac{\lambda K}{\gamma_c^\lambda \tilde{q}_{*c}^\lambda} \to \infty$. This produces a contradiction to (54). Therefore, we have the first equation of (A-23).

Next we show the second equation of (A-23). Suppose contrariwise that there exists a sequence of $\{\lambda_k, k \geq 1\}$ such that

$$\lim_{k\to\infty} \frac{\tilde{q}_{*c}^{\lambda_k}}{\sqrt{\lambda_k}} = \infty. \tag{A-28}$$

Again, for simpler notation, we label the sequence by λ. From (52) and the strict convexity of $C(\cdot)$, in view of (A-28), we know that

$$\lim_{\lambda\to\infty} \frac{\alpha_c(\tilde{q}_{*c}^\lambda)\cdot\tilde{q}_{*c}^\lambda}{\sqrt{\lambda}} = \infty, \quad \lim_{\lambda\to\infty} \frac{[1-\alpha_c(\tilde{q}_{*c}^\lambda)]\cdot\tilde{q}_{*c}^\lambda}{\sqrt{\lambda}} = \infty. \tag{A-29}$$

It follows from (A-29) and the strict convexity of $C(\cdot)$ that

$$\begin{aligned} &C(z_* - \kappa_c(\tilde{q}_{*c}^\lambda)) - \frac{\gamma_c^\lambda}{\tilde{q}_{*c}^\lambda} \int_{z_*-\kappa_c(\tilde{q}_{*c}^\lambda)}^{z_*+\overline{\kappa}_c(\tilde{q}_{*c}^\lambda)} C(x)\mathsf{d}x \\ &= \lim_{\lambda\to\infty} \frac{\gamma_c^\lambda}{\tilde{q}_{*c}^\lambda} \int_{z_*-\kappa_c(\tilde{q}_{*c}^\lambda)}^{z_*+\overline{\kappa}_c(\tilde{q}_{*c}^\lambda)} \Big(C(z_* - \kappa_c(\tilde{q}_{*c}^\lambda)) - C(x)\Big)\mathsf{d}x \\ &> 0. \end{aligned}$$

But, from (A-28), $\lim_{\lambda\to\infty} \lambda K/(\gamma_c^\lambda \tilde{q}_{*c}^\lambda) = 0$. Thus, we reach a contradiction to (54). In other words, (A-28) cannot hold, and we must have the second equation of (A-23).

As $\{\alpha_c(\tilde{q}_{*c}^\lambda) \;:\; \lambda > 0\}$ is also bounded, in view of (A-23), we pick up any two convergence sequences, say $\{\frac{\tilde{q}_{*c}^{\lambda_k}}{\sqrt{\lambda_k}} \;:\; k \geq 1\}$ and $\{\alpha_c(\tilde{q}_{*c}^{\lambda_k}) \;:\; k \geq 1\}$, from $\{\frac{\tilde{q}_{*c}^\lambda}{\sqrt{\lambda}} \;:\; \lambda > 0\}$ and $\{\alpha_c(\tilde{q}_{*c}^\lambda) \;:\; \lambda > 0\}$ (again, write them as λ sequences). Let

$$\lim_{k\to\infty} \frac{\tilde{q}_{*c}^\lambda}{\sqrt{\lambda}} = \overline{\tau} \quad \text{and} \quad \lim_{k\to\infty} \alpha_c(\tilde{q}_{*c}^\lambda) = \overline{\alpha}. \tag{A-30}$$

These imply

$$\lim_{\lambda\to\infty} \Big(\kappa_c(\tilde{q}_{*c}^\lambda) - \alpha\tau\eta\Big) = \lim_{\lambda\to\infty} \Big(\overline{\kappa}_c(\tilde{q}_{*c}^\lambda) - (1-\alpha)\tau\eta\Big) = 0.$$

We have, by (52), that

$$\lim_{\lambda\to\infty} \frac{\gamma_c^\lambda}{\tilde{q}_{*c}^\lambda} \int_{z_*-\kappa_c(\tilde{q}_{*c}^\lambda)}^{z_*+\overline{\kappa}_c(\tilde{q}_{*c}^\lambda)} C(x)\mathsf{d}x = \frac{1}{\overline{\tau}\eta} \int_{z_*-\overline{\alpha}\overline{\tau}\eta}^{z_*+(1-\overline{\alpha})\overline{\tau}\eta} C(y)\mathsf{d}y, \quad \text{(A-31)}$$

$$\lim_{\lambda\to\infty} \frac{\lambda K}{\gamma_c^\lambda \tilde{q}_{*c}^\lambda} = \frac{K\eta}{\overline{\tau}}, \quad \text{(A-32)}$$

$$C\left(z_* - \overline{\alpha}\overline{\tau}\eta\right) = C\left(z_* + (1-\overline{\alpha})\overline{\tau}\eta\right). \quad \text{(A-33)}$$

It follows from (54), (A-31)-(A-32) that

$$\frac{K\eta}{\overline{\tau}} = C\left(z_* - \overline{\alpha}\overline{\tau}\eta\right) - \frac{1}{\overline{\tau}\eta} \int_{z_*-\overline{\alpha}\overline{\tau}\eta}^{z_*+(1-\overline{\alpha})\overline{\tau}\eta} C(y)\mathsf{d}y. \quad \text{(A-34)}$$

Thus the limits of any convergence sequences of $\{\frac{\tilde{q}_{*c}^\lambda}{\sqrt{\lambda}}\}$ and $\{\alpha(\tilde{q}_{*c}^\lambda)\}$ satisfy (A-33)-(A-34). By Lemma 6, we proved (B.i) and (B.ii) for $(\tilde{r}_{*c}^\lambda, \tilde{q}_{*c}^\lambda)$ of system-$\tilde{\mathcal{S}}_c^\lambda$.

Now consider (B.iii) for $\widetilde{\mathcal{AC}}_c(\tilde{r}_{*c}^\lambda, \tilde{q}_{*c}^\lambda)$. Similar to (63), using the Taylor expansion, we get

$$\begin{aligned} \int_{z_*-\kappa_c(\tilde{q}_{*c}^\lambda)}^{z_*+\overline{\kappa}_c(\tilde{q}_{*c}^\lambda)} C(y)\mathsf{d}y &= \int_{z_*-\kappa_c(\tilde{q}_{*c}^\lambda)}^{0} C(y)\mathsf{d}y + \int_0^{z_*+\overline{\kappa}_c(\tilde{q}_{*c}^\lambda)} C(y)\mathsf{d}y \\ &= \int_{z_*-\alpha\tau\eta}^{0} C(y)\mathsf{d}y + C\left(z_* - \alpha\tau\eta\right) \times \left(\kappa_c(\tilde{q}_{*c}^\lambda) - \alpha\tau\eta\right) + \int_0^{z_*+(1-\alpha)\tau\eta} C(y)\mathsf{d}y \\ &\quad + C\left(z_* + (1-\alpha)\tau\eta\right) \times \left(\overline{\kappa}_c(\tilde{q}_{*c}^\lambda) - (1-\alpha)\tau\eta\right) + o(1). \end{aligned} \quad \text{(A-35)}$$

This, by the first part of the proposition and (A-35), implies that

$$\begin{aligned} \widetilde{\mathcal{AC}}_c(\tilde{r}_{*c}^\lambda, \tilde{q}_{*c}^\lambda) &= \frac{\lambda K}{\tilde{q}_{*c}^\lambda} + \gamma_c^\lambda \cdot \frac{\gamma_c^\lambda}{\tilde{q}_{*c}^\lambda} \int_{z_*-\kappa_c(\tilde{q}_{*c}^\lambda)}^{z_*+\overline{\kappa}_c(\tilde{q}_{*c}^\lambda)} C(y)\mathsf{d}y \\ &= \left(\frac{K}{\tau} + \frac{1}{\tau\eta^2} \int_{z_*-\alpha\tau\eta}^{z_*+(1-\alpha)\tau\eta} C(y)\mathsf{d}y\right)\sqrt{\lambda} + o(\sqrt{\lambda}). \end{aligned} \quad \text{(A-36)}$$

Therefore, (B.iii) for $\widetilde{\mathcal{AC}}_c(\tilde{r}_{*c}^\lambda, \tilde{q}_{*c}^\lambda)$ is proved.

Now consider $K = 0$. $\tilde{q}_{*c}^\lambda = 1$ directly follows from (48) and convexity of $C(\cdot)$. (B.ii) and (B.iii) are given by (45) and (48). ■

Proof : **[of Proposition 6]** We first prove (i). Suppose that

$$\overline{\lim_{\lambda\to\infty}} \Delta_c(\lambda) = \infty. \quad \text{(A-37)}$$

Then, by (48),

$$\begin{aligned} \lim_{\lambda\to\infty} \frac{\min_{r^\lambda} \widetilde{\mathcal{AC}}_c(r^\lambda, \tilde{q}^\lambda)}{\sqrt{\lambda}} &= \lim_{\lambda\to\infty} \frac{\gamma_c^\lambda}{\tilde{q}^\lambda\sqrt{\lambda}} \int_{\tilde{r}_{*c}^\lambda(\tilde{q}^\lambda)}^{\tilde{r}_{*c}^\lambda(\tilde{q}^\lambda)+\tilde{q}^\lambda} C\left(z_c^\lambda(y)\right)\mathsf{d}y \\ &= \frac{(\gamma_c^\lambda)^2}{\sqrt{\lambda}\tilde{q}^\lambda} \int_{z_c^\lambda(\tilde{r}_{*c}^\lambda(\tilde{q}^\lambda))}^{z_c^\lambda(\tilde{r}_{*c}^\lambda(\tilde{q}^\lambda))+\beta_c^\lambda(\tilde{q}^\lambda)} C(y)\ \mathsf{d}y. \end{aligned} \quad \text{(A-38)}$$

Applying L'Hôpital's rule, we have

$$\lim_{\lambda\to\infty}\int_{z_c^\lambda(\tilde{r}_{*c}^\lambda(\tilde{q}^\lambda))}^{z_c^\lambda(\tilde{r}_{*c}^\lambda(\tilde{q}^\lambda))+\beta_c^\lambda(\tilde{q}^\lambda)} C(y)\ \mathsf{d}y \Big/ \frac{\tilde{q}^\lambda}{\sqrt{\lambda}} = \infty. \tag{A-39}$$

Hence, by (A-38)-(A-39) and Proposition 5, we have that

$$\overline{\lim_{\lambda\to\infty}} \frac{\min_{r^\lambda} \widetilde{\mathcal{AC}}_c(r^\lambda, \tilde{q}^\lambda)}{\widetilde{\mathcal{AC}}_c(\tilde{r}_{*c}^\lambda, \tilde{q}_{*c}^\lambda)} = \infty,$$

which implies (i). Now suppose that $\overline{\lim}_{\lambda\to\infty}\Delta_c(\lambda) < \infty$ but $\overline{\lim}_{\lambda\to\infty}\Delta_c(\lambda) \neq \underline{\lim}_{\lambda\to\infty}\Delta_c(\lambda)$. Then there exist two convergence sequences, say $\{\frac{\tilde{q}^{\lambda_k}}{\sqrt{\lambda_k}} \ : \ k \geq 1\}$ and $\{\alpha_c(\tilde{q}^{\lambda_k}) \ : \ k \geq 1\}$, from $\{\frac{\tilde{q}^\lambda}{\sqrt{\lambda}} \ : \ \lambda > 0\}$ and $\{\alpha_c(\tilde{q}^\lambda) \ : \ \lambda > 0\}$ (again, write them as λ sequences) such that

$$\lim_{k\to\infty}\frac{\tilde{q}^\lambda}{\sqrt{\lambda}} = \tilde{\tau} \neq \tau \quad \text{and} \quad \lim_{k\to\infty}\alpha_c(\tilde{q}^\lambda) = \tilde{\alpha} \neq \alpha. \tag{A-40}$$

Exactly going along the line (A-35)-(A-36), we have

$$\widetilde{\mathcal{AC}}_c(\tilde{r}_{*c}^\lambda(\tilde{q}^\lambda), \tilde{q}^\lambda) \quad = \quad \left(\frac{K}{\tilde{\tau}} + \frac{1}{\tilde{\tau}\eta^2}\int_{z_*-\tilde{\alpha}\tilde{\tau}\eta}^{z_*+(1-\tilde{\alpha})\tilde{\tau}\eta} C(y)\mathsf{d}y\right)\sqrt{\lambda} + o(\sqrt{\lambda}), \tag{A-41}$$

where $\tilde{\alpha}$, using (52), satisfies

$$C\left(z_* - \tilde{\alpha}\tilde{\tau}\eta\right) = C\left(z_* + (1-\tilde{\alpha})\tilde{\tau}\eta\right). \tag{A-42}$$

Consider function

$$g_c(\tilde{\tau}) = \frac{K}{\tilde{\tau}} + \frac{1}{\tilde{\tau}\eta^2}\int_{z_*-\tilde{\alpha}\tilde{\tau}\eta}^{z_*+(1-\tilde{\alpha})\tilde{\tau}\eta} C(y)\mathsf{d}y.$$

If $-g_c(\tilde{\tau})$ is strictly unimodal and its maximizer is given by τ, then we have (i). Thus to complete the proof of (i), it is sufficient to show that the strict unimodality of $-g_c(\tilde{\tau})$ and its maximizer is τ. Note, by (A-42), that

$$\frac{\mathsf{d}g_c(\tilde{\tau})}{\mathsf{d}\tilde{\tau}} = -\frac{K}{\tilde{\tau}^2} - \frac{1}{\tilde{\tau}^2\eta^2}\int_{z_*-\tilde{\alpha}\tilde{\tau}\eta}^{z_*+(1-\tilde{\alpha})\tilde{\tau}\eta} C(y)\mathsf{d}y + \frac{1}{\tilde{\tau}\eta}C(z_* - \tilde{\alpha}\tilde{\tau}\eta).$$

Letting $\mathsf{d}g_c(\tilde{\tau})/\mathsf{d}\tilde{\tau} = 0$, we have

$$\eta^2 K = \tilde{\tau}\eta \times C\left(z_* - \tilde{\alpha}\tilde{\tau}\eta\right) - \int_{z_*-\tilde{\alpha}\tilde{\tau}\eta}^{z_*+(1-\tilde{\alpha})\tilde{\tau}\eta} C(y)\mathsf{d}y. \tag{A-43}$$

Making a comparison with (55), we know τ is minimizer of $g(\tilde{\tau})$. Considering

$$\tilde{\tau}\eta \times C\left(z_* - \tilde{\alpha}\tilde{\tau}\eta\right) - \int_{z_*-\tilde{\alpha}\tilde{\tau}\eta}^{z_*+(1-\tilde{\alpha})\tilde{\tau}\eta} C(y)\mathsf{d}y$$

as a function of $\tilde{\tau}\eta$, by (A-42) and Lemma 6 in Zheng (1992), it is strict increasing. Hence, we know that

$$-\frac{K}{\tilde{\tau}^2} - \frac{1}{\tilde{\tau}^2\eta^2}\int_{z_*-\tilde{\alpha}\tilde{\tau}\eta}^{z_*+(1-\tilde{\alpha})\tilde{\tau}\eta} C(y)\mathsf{d}y + \frac{1}{\tilde{\tau}\eta}C(z_* - \tilde{\alpha}\tilde{\tau}\eta) < 0 \quad \text{for } \tilde{\tau} < \tau;$$
$$-\frac{K}{\tilde{\tau}^2} - \frac{1}{\tilde{\tau}^2\eta^2}\int_{z_*-\tilde{\alpha}\tilde{\tau}\eta}^{z_*+(1-\tilde{\alpha})\tilde{\tau}\eta} C(y)\mathsf{d}y + \frac{1}{\tilde{\tau}\eta}C(z_* - \tilde{\alpha}\tilde{\tau}\eta) > 0 \quad \text{for } \tilde{\tau} > \tau.$$

Thus the unimodality of $-g_c(\tilde{\tau})$ is proven.

Finally we prove (ii). Suppose that $\overline{\lim}_{\lambda\to\infty} |\varpi^\lambda| < \infty$ and one of $\overline{\lim}_{\lambda\to\infty} \varpi^\lambda \neq 1$ and $\underline{\lim}_{\lambda\to\infty} \varpi^\lambda \neq 1$ holds. Then there exists a convergence sequences, say $\{\varpi^{\lambda_k} \ : \ k \geq 1\}$ from $\{\varpi^\lambda \ : \ \lambda > 0\}$ (again, write them as λ sequences) such that

$$\lim_{\lambda\to\infty} \varpi^\lambda = b \neq 1. \tag{A-44}$$

Similar to (A-36),

$$\lim_{\lambda\to\infty} \frac{\widetilde{\mathcal{AC}}_c(\tilde{r}^\lambda, \tilde{q}_{*c}^\lambda)}{\sqrt{\lambda}} = \frac{K}{\tau} + \frac{1}{\tau\eta^2}\int_{b(z_*-\alpha\tau\eta)}^{b(z_*-\alpha\tau\eta)+\tau\eta} C(y)\mathsf{d}y.$$

It is direct to verify the above function has a unique minimizer at $b = 1$. Hence we have (ii). ■

Proof : [**of Proposition 7**] By (10)-(11) and (46), going along the line of the proof of Propositions 3 and 4, we can show the proposition holds. Here the details are omitted. ■

Proof : [**of Equation (107)**] Suppose contrariwise that

$$\overline{\lim_{\lambda\to\infty}} \sqrt{\lambda}\hat{T}_*^\lambda = \infty.$$

Then there exists a subsequence $\{\lambda_k : k \geq 1\}$ (label it as λ sequence) such that

$$\lim_{k\to\infty} \sqrt{\lambda}\hat{T}_*^\lambda = \infty. \tag{A-45}$$

By (103) and (106),

$$\lim_{\lambda\to\infty} \left[C\left(\frac{z_*\theta\sqrt{\rho^\lambda} + M^\lambda(\hat{T}_*^\lambda) - \lambda\hat{T}_*^\lambda}{\theta\sqrt{\lambda\hat{T}_*^\lambda + \rho^\lambda}} \right) \sqrt{\hat{T}_*^\lambda + \frac{1}{\mu}} \right.$$
$$\left. - \frac{1}{\hat{T}_*^\lambda}\int_0^{\hat{T}_*^\lambda} C\left(\frac{z_*\theta\sqrt{\rho^\lambda} + M^\lambda(\hat{T}_*^\lambda) - \lambda t}{\theta\sqrt{\lambda t + \rho^\lambda}} \right) \sqrt{t + \frac{1}{\mu}}\mathsf{d}t \right] = 0. \tag{A-46}$$

The remainder of the proof is divided into three cases.

Case A $\lim_{\lambda\to\infty} \hat{T}_*^\lambda = 0$; **Case B** $\lim_{\lambda\to\infty} \hat{T}_*^\lambda = a \in (0, \infty)$; **Case C** $\lim_{\lambda\to\infty} \hat{T}_*^\lambda = \infty$.

For each case, we will get a contradiction with (A-46) if (A-45) holds. First we look at Case A. This case will be further divided into subcases by (103) and (A-45):

Subcase A.1 $\lim_{\lambda\to\infty}\hat{T}_*^\lambda = 0$ and $\lim_{\lambda\to\infty}\frac{M^\lambda(\hat{T}_*^\lambda)-\lambda\hat{T}_*^\lambda}{\theta\sqrt{\lambda\hat{T}_*^\lambda+\rho^\lambda}} = -\infty$.

Under Subcase A.1, we have

$$\lim_{\lambda\to\infty}\frac{z_*\theta\sqrt{\rho^\lambda}+M^\lambda(\hat{T}_*^\lambda)-\lambda\hat{T}_*^\lambda}{\theta\sqrt{\lambda\hat{T}_*^\lambda+\rho^\lambda}} = -\infty \quad\text{and}\quad \lim_{\lambda\to\infty}\frac{z_*\theta\sqrt{\rho^\lambda}+M^\lambda(\hat{T}_*^\lambda)}{\theta\sqrt{\rho^\lambda}} \geq z_*.$$

Then it follows from the strict convexity of $C(\cdot)$ that

$$\lim_{\lambda\to\infty}\left|C\left(\frac{z_*\theta\sqrt{\rho^\lambda}+M^\lambda(\hat{T}_*^\lambda)-\lambda\hat{T}_*^\lambda}{\theta\sqrt{\lambda\hat{T}_*^\lambda+\rho^\lambda}}\right)-\frac{1}{\hat{T}_*^\lambda}\int_0^{\hat{T}_*^\lambda}C\left(\frac{z_*\theta\sqrt{\rho^\lambda}+M^\lambda(\hat{T}_*^\lambda)-\lambda t}{\theta\sqrt{\lambda t+\rho^\lambda}}\right)\mathsf{d}t\right| = \infty,$$

this, in view of $\lim_{\lambda\to\infty}\hat{T}_*^\lambda = 0$, contradicts with (A-46).

Subcase A.2 $\lim_{\lambda\to\infty}\hat{T}_*^\lambda = 0$ and $\lim_{\lambda\to\infty}\frac{M^\lambda(\hat{T}_*^\lambda)-\lambda\hat{T}_*^\lambda}{\theta\sqrt{\lambda\hat{T}_*^\lambda+\rho^\lambda}} = b$ with $|b| < \infty$.

For this subcase, by (A-45), we have

$$\lim_{\lambda\to\infty}\frac{z_*\theta\sqrt{\rho^\lambda}+M^\lambda(\hat{T}_*^\lambda)-\lambda\hat{T}_*^\lambda}{\theta\sqrt{\lambda\hat{T}_*^\lambda+\rho^\lambda}} = z_*+b \quad\text{and}\quad \lim_{\lambda\to\infty}\frac{z_*\theta\sqrt{\rho^\lambda}+M^\lambda(\hat{T}_*^\lambda)}{\theta\sqrt{\rho^\lambda}} = \infty.$$

Similar to Subcase A.1, by the strict convexity of $C(\cdot)$, we get a contradiction with (A-46). Cases B and C can be analyzed along the same line. ■