MATHEMATICS OF OPERATIONS RESEARCH
Vol. 22, No. 1, February 1997
Printed in U.S.A.

CORRECTED DIFFUSION APPROXIMATIONS FOR A MULTISTAGE PRODUCTION-INVENTORY SYSTEM

PAUL GLASSERMAN AND TAI-WEN LIU

We analyze a multistage inventory system with limited production capacity facing stochastic demands. Each node follows a periodic-review base-stock policy for echelon inventory: in each period, each node attempts to produce enough material to restore cumulative downstream inventory to a fixed target level. We develop approximations to the key measures of interest (average inventories, average backorders, and service levels) by simultaneously letting the mean demand approach the system's bottleneck capacity and letting the base-stock level for finished goods increase without bound. Using a method of Siegmund, we thus obtain diffusion limits with higher-order correction terms. A numerical example suggests that the correction terms can substantially improve the accuracy of the approximations.

1. Introduction and main results. Among the most fundamental models in multiechelon inventory theory is the facilities-in-series model of Clark and Scarf (1960). In this system, the top node draws raw material from an external source; each intermediate node orders material from its predecessor and supplies its successor; and the bottom node supplies external demands. Optimal ordering decisions follow an *echelon base-stock policy*, in which each node orders just enough in each period to restore its cumulative downstream inventory position to a fixed target level. The term *echelon* indicates that ordering decisions are tied to cumulative inventories, and the *base-stock* level is the target to which echelon inventory is to be restored.

The Clark-Scarf model places no limit on the amount of material that can move through a facility in a single period. To model processing or production activity at a node explicitly, it is generally necessary to assign a capacity to the node which then specifies an upper limit on the inventory that can move through the node in a period. If we let d denote the number of nodes; c^i, $i = 1, \ldots, d$, their capacities; s^i, $i = 1, \ldots, d$, the echelon base-stock levels; and D_n the total demand in period n, $n \geq 1$, then the dynamics of the capacity-constrained system are fully described by the following recursions:

$$(1) \quad Y_n^i = \max\{0, Y_{n-1}^i + D_n - c^i, Y_{n-1}^{i+1} + D_n - (s^{i+1} - s^i)\}; \qquad i = 1, \ldots, d-1;$$

$$(2) \quad Y_n^d = \max\{0, Y_{n-1}^d + D_n - c^d\}.$$

The *shortfall* Y_n^i records the difference between the target level s^i and the actual inventory for echelon i in period n; thus, $s^i - Y_n^i$ is the cumulative net inventory in nodes $1, \ldots, i$. Under a base-stock policy, each node i attempts to order and process sufficient material in each period to drive its shortfall to zero, while not exceeding its own production capacity or the available upstream inventory. The second and third expressions inside the max in (1) reflect the capacity and inventory constraints,

Received October 28, 1994; revised February 28, 1996 and April 12, 1996.
AMS 1991 subject classification. Primary: 90B20; Secondary: 60J60.
OR/MS Index 1978 subject classification. Primary: Inventory/Production/Multistage; Secondary: Probability/Diffusion.
Key words. Diffusion approximation, heavy traffic, perturbed random walks, Wald's identity.

0364-765X/97/2201/0186/$05.00

respectively. The inventory constraint is absent in (2) because node d draws raw material from an unlimited external source. For a more detailed derivation of (1)–(2) and further discussion of modeling and applications, see Glasserman and Tayur (1994, 1996) and references there.

No optimal policy is known for the capacity-constrained Clark-Scarf model, but the base-stock policy reflected in (1)–(2) remains attractive because of its simplicity, because it is optimal in the unconstrained case, and because it remains optimal in a single-node capacity-constrained system, as shown by Federgruen and Zipkin (1986). For recent work on the control of multistage capacity-constrained systems and base-stock policies in particular, see Schraner (1995) and Speck and van der Wal (1991a, b); for a survey of work on multistage systems generally, see van Houtman, Inderfurth, and Zijm (1995). Even if we restrict attention to base-stock policies, evaluating performance under a particular set of policy parameters is difficult; our objective is to present accurate approximations to the key measures of performance.

If the demands $\{D_n,\ n \geq 1\}$ are i.i.d. and have mean less than

$$c^* = \min\{c^1, \ldots, c^d\},$$

then the shortfalls defined in (1)–(2) converge to a finite stationary distribution from all initial values. Let $(Y^1, \ldots, Y^d)$ have this stationary distribution. The quantities we consider are the mean shortfalls $\mathsf{E}[Y^i]$, $i = 1, \ldots, d$; the *stockout probability* $P(Y^1 > s^1)$; the *average backlog* $\mathsf{E}(Y^1 - s^1)^+$; and the *unfilled demand* $u(s^1) = \mathsf{E}(\min\{Y^1 + D - s^1, D\})^+$. When linear costs are charged on inventories and backorders, the mean shortfalls and the average backlog can be combined to give the average cost per period; hence, approximations for these quantities provide approximations to average linear costs as well. It follows from (1)–(2) that approximating $\mathsf{E}[Y^i]$ for $i > 1$ is a special case of approximating $\mathsf{E}[Y^1]$, so we consider only the latter explicitly. The unfilled demand is primarily of interest in defining the *fill rate*

$$f(s^1) = 1 - \frac{u(s^1)}{\mathsf{E}[D]} = 1 - \frac{\mathsf{E}(\min\{Y^1 + D - s^1, D\})^+}{\mathsf{E}[D]}, \tag{3}$$

usually considered the key measure of service.

We develop approximations to these quantities as s^1 becomes large, $\Delta^i = s^{i+1} - s^i$, $i = 1, \ldots, d-1$, remain fixed, and the mean demand approaches c^*, based on the method of Siegmund (1979). (See Asmussen 1984, Chang 1992, and Hogan 1986 for further development and application of this method.) We assume the common distribution of the random variables $X_n \triangleq D_n - c^*$, $n \geq 1$, is a member of an exponential family $\{F_\theta,\ \theta \in \Theta\}$; i.e., a family of distributions admitting the representation

$$dF_\theta(x) = \exp\{\theta x - \psi(\theta)\}\, dF_0(x)$$

for some distribution F_0 with support in $[-c^*, \infty)$ and cumulant generating function $\psi(\theta) = \log \mathsf{E}_0[\exp\{\theta(D - c^*)\}]$ assumed finite in a neighborhood of the origin. (This is equivalent to assuming the demand distribution is from an exponential family, but it is more convenient to impose the conditions on the X_n.) We always have $\psi(0) = 0$, and without loss of generality we adopt the normalization $\psi'(0) = 0$, $\psi''(0) = 1$. It is easy to see that $\psi'(\theta) = \mu_\theta \triangleq \mathsf{E}_\theta[X_1]$, $\psi''(\theta) = \mathrm{Var}_\theta[X_1]$, and that $\mathsf{E}_\theta[X_1]$ and θ have the same sign. A Taylor expansion of ψ about $\theta = 0$ yields

$$\psi(\theta) = \tfrac{1}{2}\theta^2 + o(\theta^2), \quad \text{as } \theta \to 0, \tag{4}$$

and therefore

$$\mu_\theta = \psi'(\theta) = \theta + o(\theta), \quad \text{as } \theta \to 0. \tag{5}$$

Furthermore, ψ is strictly convex wherever it is finite, so for each sufficiently small $\theta_0 < 0$ there is just one $\theta_1 > 0$ for which $\psi(\theta_0) = \psi(\theta_1)$. We set $\gamma = \theta_1 - \theta_0$ and note that the condition $\gamma \to 0$ is equivalent to $\theta_0 \to 0$, $\mu_{\theta_0} \to 0$, and thus to $\mathsf{E}_{\theta_0} D \to c^*$.

Our approximations are sharpest when the distribution F_0 is *strongly nonlattice*, meaning that the characteristic function $g(\lambda) = \mathsf{E}_0[\exp(i\lambda X)]$ satisfies $\inf_{|\lambda| > \delta} |1 - g(\lambda)| > 0$ for each $\delta > 0$. This is equivalent to assuming that the demand distribution itself is strongly nonlattice. A strongly nonlattice distribution is indeed nonlattice; all spread-out distributions are strongly nonlattice (Asmussen 1984, p. 142).

We need some additional notation to state our main result. Let $S_n = \sum_{i=1}^n X_i$ and let

$$\tau_+ = \inf\{n \geq 1 \colon S_n > 0\}$$

be the first strong ascending ladder epoch for this random walk. Let $\beta = \mathsf{E}_0[S_{\tau_+}^2]/(2\mathsf{E}_0[S_{\tau_+}])$ and $\kappa = \mathsf{E}_0[S_{\tau_+}^3]/(3\mathsf{E}_0[S_{\tau_+}])$. Finally, let $j^* = \min\{1 \leq i \leq d \colon c^i = c^*\}$ be the index of the lowest bottleneck and define

$$\xi = \max_{i \geq j^*} \left\{ (i-1)c^* - \sum_{k=1}^{i-1} \Delta^k \right\}.$$

We now have

THEOREM 1. *Suppose that F_0 is strongly nonlattice and that $\theta_0 \uparrow 0$, $b \to \infty$ in such a way that $\theta_0 b \to$ constant. Then*

(i) *the mean shortfall at node 1 satisfies* $\mathsf{E}_{\theta_0} Y^1 = \gamma^{-1} e^{-\gamma(\beta - \xi)} + O(\gamma)$;

(ii) *the stockout probability satisfies* $P_{\theta_0}\{Y^1 > b\} = e^{-\gamma(b + \beta - \xi)} + o(\gamma^2)$;

(iii) *the average backlog satisfies* $\mathsf{E}_{\theta_0}(Y^1 - b)^+ = \gamma^{-1} e^{-\gamma(b + \beta - \xi)} + o(\gamma)$;

(iv) *the unfilled demand satisfies* $u(b) = \gamma^{-1} e^{-\gamma(b + \beta - \xi)}(e^{\gamma c^*} - 1) + o(\gamma^{2-\epsilon})$, *for all* $\epsilon > 0$.

If F_0 is merely assumed nonlattice, the error terms in (ii), (iii) *and* (iv) *become* $o(\gamma)$, $o(1)$, *and* $o(\gamma^{1-\epsilon})$ *respectively*; (i) *is unchanged.*

REMARKS. (a) Siegmund (1979, p. 716), and Siegmund (1985, p. 225) give an integral representation of β suitable for numerical evaluation. Asmussen's (1992) results suggest a matrix-analytic approach to computation of β for phase-type demands. Since γ is easily computed as the root of an equation, it follows that the expressions in the theorem can be evaluated with minimal computational effort.

(b) In a single-node system we have $\xi = 0$ and directly from Theorem 1 of Siegmund (1979) we get the finer approximation

$$\mathsf{E}_{\theta_0} Y^1 = \frac{1}{\gamma} - \beta + \frac{\gamma}{2}[\kappa - \beta^2] + o(\gamma), \tag{6}$$

for nonlattice F_0. Part (ii) becomes

$$P_{\theta_0}\{Y^1 > b\} = e^{-\gamma(b+\beta)} + o(\gamma^2), \tag{7}$$

assuming a strongly nonlattice F_0, which coincides with Theorem 2 of Siegmund

(1979). Parts (iii) and (iv) are new even for single-stage systems, and in this case (iv) can be strengthened to

$$u(b) = \frac{1}{\gamma} e^{-\gamma(b+\beta)}(e^{\gamma c^*} - 1) + o(\gamma^2). \tag{8}$$

(c) The fill rate in (3) can be approximated using part (iv) of the theorem and the mean demand, which is presumably known. Alternatively, we can use (5) to get

$$\mathsf{E}_{\theta_0} D = \mathsf{E}_{\theta_0} X + c^* = c^* + \theta_0 + o(\theta_0), \quad \text{as } \theta_0 \uparrow 0,$$

and substitute this in (3) to get

$$1 - f(b) = \left(\frac{1}{\gamma c^*} + \frac{1}{2c^{*2}} \right) e^{-\gamma(b+\beta-\xi)}(e^{\gamma c^*} - 1) + o(\gamma), \tag{9}$$

for strongly nonlattice demands.

(d) The case of lattice F_0 requires care but leads to similar results. A detailed analysis is given in Liu (1995); we summarize its conclusions: The approximations in (i)–(iii) are unchanged; except that β is replaced with $\beta + l/2$ in (ii), where l is the span of F_0. The error terms for (i)–(iii) are $O(\gamma)$, $o(\gamma)$, and $o(1)$, respectively. The approximation in (iv) becomes $\gamma^{-1} \exp\{-\gamma(b + \beta - \xi)\}(\exp\{\gamma l[c^*/l]\} - 1)$, where $[\cdot]$ denotes the integer part; the corresponding error term becomes $o(1)$. An assumption in the lattice case is that $b - \xi$ increases through multiples of l.

Table 1 compares the corrected diffusion approximations in parts (i) and (ii) of Theorem 1 with ordinary diffusion approximations. These numerical results are for a two-node system with $c^1 = 2$ and $c^2 = c^* = 1$; three values of $\Delta = s^2 - s^1$; and two values of $\rho \triangleq \mathsf{E}[D]/c^*$. Demands are exponentially distributed, so $\beta = c^*$, and the exact distribution of Y^1 can also be found explicitly. The ordinary Brownian approximations to $\mathsf{E}Y^1$ and $P(Y^1 > x)$ are $\sigma^2/(2|\mu|)$ and $\exp(-2|\mu|x/\sigma^2)$, where $\mu = \mathsf{E}D - c^*$ and σ^2 is the demand variance. These approximations are thus insensitive to Δ. For $0 \le \Delta \le c^1$, the corrected approximation is exact in this example. The results in the table suggest significant improvements from the correction terms, especially at moderate ρ but even at very high ρ. Of course, the corrected approximations rely on stronger independence assumptions and more detailed distributional information than the ordinary Brownian approximations.

Comparing corrected approximations with Brownian limits, Siegmund (1979) interprets the approximation in (7) as follows: using γ instead of $2|\mu|/\sigma^2$ corrects for non-normality of the X_n; adding β to the boundary b corrects for discontinuity and

TABLE 1
Comparison of Exact Values, Corrected Approximations, and Ordinary Brownian Approximations for a Two-node System with Exponentially Distributed Demands

		EY^1			$P(Y^1 > 3)$		
	Δ	Exact	Theorem 1	Brownian	Exact	Theorem 1	Brownian
$\rho = 0.60$	1.5	0.1639	0.1639	0.4500	0.00629	0.00629	0.00127
	2.25	0.0757	0.0704	0.4500	0.00276	0.00270	0.00127
	2.5	0.0624	0.0532	0.4500	0.00214	0.00204	0.00127
$\rho = 0.98$	1.5	23.206	23.206	24.010	0.8332	0.8332	0.8825
	2.25	22.512	22.510	24.010	0.8083	0.8082	0.8825
	2.5	22.286	22.283	24.010	0.8002	0.8001	0.8825

further accounts for the distribution of the X_n. To this interpretation we add that, in Theorem 1, the term ξ corrects for the difference between single- and multi-node systems, a distinction that vanishes in the Brownian limit.

As an application of Theorem 1, we approximate the finished-goods base-stock level required to meet a service-level constraint; i.e., for fixed $0 < \delta < 1$, we pick s_δ^1 so that either the fraction of periods without a stockout or the fill rate is approximately $1 - \delta$.

COROLLARY 1. *Suppose F_0 is strongly nonlattice and fix $0 < \delta < 1$.*
(i) *If*

$$s_\delta^1 = -\frac{1}{\gamma}\log \delta - \beta + \xi,$$

then $P_{\theta_0}(Y^1 > s_\delta^1) = \delta + o(\gamma^2)$.
(ii) *If*

$$s_\delta^1 = \begin{cases} -\dfrac{1}{\gamma}\log \delta - \beta + \xi + \dfrac{c^*}{2} + \dfrac{1}{\gamma}\log\dfrac{c^*}{\mathsf{E}_{\theta_0} D}, & \text{or} \\ -\dfrac{1}{\gamma}\log \delta - \beta + \xi + \dfrac{c^*}{2} + \dfrac{1}{2c^*}, \end{cases}$$

then $1 - f(s_\delta^1) = \delta + o(\gamma)$.

The subsequent sections of this article are devoted to proving the results above. We conclude this introduction with a general description of the analysis. A starting point is the representation, derived in Glasserman (1993),

$$Y^1 \stackrel{\mathrm{d}}{=} \max_{n \geq 0}\{S_n + \xi_n\}, \tag{10}$$

where

$$\xi_n = nc^* - r_n, \tag{11}$$

$r_0 = 0$, and for all $n \geq 1$, r_n is the length of the shortest n-step path through the graph in Figure 1, starting from the lower-left corner. It follows that there is a finite

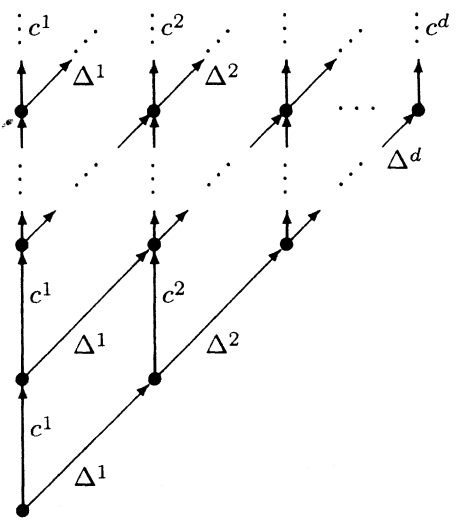

FIGURE 1. Each vertical arc in column i has length c^i, each diagonal arc from column i to column $i + 1$ has length Δ^i.

n^* such that

$$\xi_n = \xi \quad \text{for all } n \geq n^*, \tag{12}$$

and our results apply in any setting of the type in (10) if (12) holds. Indeed, we prove an essential preliminary result in §2 under the weaker assumption that $\xi_n \to \xi$. Using (10), we relate performance measures involving Y^1 to boundary crossings of the process $Z_n = S_n + \xi_n$. (In Gut's (1992) terminology, Z_n is a *perturbed random walk*.) We approximate expectations under θ_0 by first expressing them as expectations under θ_1 using Wald's identity and a likelihood ratio. The exponential form of the likelihood ratio suggests the approximations in the theorem. The likelihood ratio is a function of S rather than Z, but because of (12) we are able to locate the random walk at boundary crossings of the perturbed process and thus carry out the approximation.

2. Preliminaries. Let $\{X_n, n \geq 1\}$ be as in §1. Set $S_n = \sum_1^n X_i$ and $Z_n = S_n + \xi_n$, $n \geq 0$, for as yet unspecified $\{\xi_n, n \geq 0\}$. For all $b > 0$ define stopping times

$$T = \inf\{n \geq 1: Z_n > b\}, \tag{13}$$

and

$$\tau' = \inf\{n \geq 1: S_n > b\}. \tag{14}$$

For $t > 0$ and $-\infty < \zeta < \infty$, let $G(t; \zeta, 1)$ denote the probability that a Brownian motion process with drift ζ and unit variance reaches 1 before time t, starting from the origin. It is well known (see, e.g., Siegmund 1979, p. 706) that if $b \to \infty$, $\mu = \mathsf{E}X_1 \to 0$ and $\mu b \to \zeta \in (-\infty, \infty)$, then

$$P_\mu\{\tau' \leq b^2 t\} \to G(t; \zeta, 1), \quad \text{for each } 0 < t < \infty. \tag{15}$$

This result extends to the perturbed process Z, in the sense that

$$P_\mu\{T \leq b^2 t\} \to G(t; \zeta, 1) \tag{16}$$

as $b \to \infty$, $\mu \to 0$ and $\mu b \to \zeta$, provided that the (possibly random) sequence $\{\xi_n\}$ satisfies

$$\sup_{0 \leq s \leq t} \left| \frac{\xi_{[ns]}}{\sqrt{n}} \right| \Rightarrow 0, \quad \text{as } n \to \infty,$$

where $\Rightarrow$ denotes convergence in distribution. This follows from the converging-together theorem, as in Theorem 4.4.6 of Chung (1974). In particular, then, (16) holds for deterministic $\{\xi_n\}$ that converge to a finite limit.

Let $R_b = S_{\tau'} - b$ denote the excess over level b for the random walk, and suppose that $\{X_n, n \geq 1\}$ have distribution F_θ. It follows from the renewal theorem that for $\theta \geq 0$ the R_b have a limit in distribution as $b \to \infty$ (through multiples of the span of X_1 in the lattice case), and for $\theta = 0$ the limit random variable R_∞ has the distribution

$$H(x) \triangleq P_0\{R_\infty \leq x\} = \frac{1}{\mathsf{E}_0 S_{\tau_+}} \int_0^x P_0\{S_{\tau_+} > y\}\, dy.$$

Lemma 3 in Siegmund (1979) shows that τ'/b^2 and $S_{\tau'} - b$ are asymptotically independent in the sense that (for nonlattice F_0),

$$P_\theta\{\tau' \le b^2 t, S_{\tau'} - b \le x\} \to G(t;\zeta,1)H(x) \tag{17}$$

as $b \to \infty$, $\theta \downarrow 0$, and $\theta b \to \zeta \in [0,\infty)$. We will need a similar result when T replaces τ':

LEMMA 1. *Suppose F_0 is nonlattice and $\{\xi_n\}$ is a sequence of numbers satisfying*

$$\xi_n \to \xi, \quad \text{as } n \to \infty. \tag{18}$$

Then for $t > 0$ and $x > 0$,

$$P_\theta\{T \le b^2 t, S_T - b \le x\} \to G(t;\zeta,1)H(x + \xi), \tag{19}$$

and for $m > 0$,

$$\mathsf{E}_\theta(S_T - b)^m \to \mathsf{E}_0(R_\infty - \xi)^m, \tag{20}$$

as $b \to \infty$, $\theta \downarrow 0$, and $\theta b \to \zeta \in [0,\infty)$.

PROOF. Observe that $\theta b \to \zeta$ implies $\mu_\theta b \to \zeta$ via (5) and define

$$\tau = \inf\{n \ge 1: S_n > b - \xi\}. \tag{21}$$

For $\epsilon > 0$ choose n_ϵ so that $\xi - \epsilon \le \xi_n \le \xi + \epsilon$ for all $n \ge n_\epsilon$. If $\theta \downarrow 0$, $b \to \infty$ and $\theta b \to \zeta$, then

$$P_\theta\{\tau < n_\epsilon\} = o(1) \tag{22}$$

and

$$P_\theta\{T < n_\epsilon\} = o(1), \tag{23}$$

by (15) and (16). Now observe that

$$P_\theta\{T \ne \tau\} \le P_\theta\{T \ne \tau, T \wedge \tau \ge n_\epsilon\} + P_\theta\{T < n_\epsilon\} + P_\theta\{\tau < n_\epsilon\}. \tag{24}$$

On the event $\{n_\epsilon \le \tau < T\}$, we have $S_\tau + \xi_\tau = Z_\tau \le b$ and $\xi - \xi_\tau \le \epsilon$, so $S_\tau \le b - \xi + \epsilon$, implying $R_{b-\xi} \le \epsilon$. Similarly, on the event $\{n_\epsilon \le T < \tau\}$, we have $R_{b-\xi-\epsilon} \le \epsilon$. Thus,

$$\begin{aligned} P_\theta\{T \ne \tau, T \wedge \tau \ge n_\epsilon\} &\le P_\theta\{R_{b-\xi} \le \epsilon\} + P_\theta\{R_{b-\xi-\epsilon} \le \epsilon\} \\ &= 2H(\epsilon) + o(1), \quad \text{as } b \to \infty, \theta \downarrow 0, b\theta \to \zeta, \end{aligned}$$

in light of (17). Since $\epsilon > 0$ is arbitrary and $H(0+) = 0$, we conclude that $P_\theta\{T \ne \tau, T \wedge \tau \ge n_\epsilon\} \to 0$, and thus by (22)–(23) that

$$P_\theta\{T \ne \tau\} = o(1), \quad \text{as } \theta \to 0, b \to \infty \text{ and } \theta b \to \zeta. \tag{25}$$

Finally,

$$\begin{aligned}
&\left|P_\theta\{T \le b^2 t, S_T - b \le x\} - G(t;\zeta,1)H(x+\xi)\right| \\
&\quad \le \left|P_\theta\{T \le b^2 t, S_T - b \le x\} - P_\theta\{\tau \le b^2 t, S_\tau - b \le x\}\right| \\
&\qquad + \left|P_\theta\{\tau \le b^2 t, S_\tau - b \le x\} - P_\theta\{\tau \le (b-\xi)^2 t, S_\tau - (b-\xi) \le x + \xi\}\right| \\
&\qquad + \left|P_\theta\{\tau \le (b-\xi)^2 t, S_\tau - (b-\xi) \le x + \xi\} - G(t;\zeta,1)H(x+\xi)\right| \\
&\quad \le P_\theta\{T \ne \tau\} + P_\theta\{(b-\xi)^2 t < \tau \le b^2 t\} + o(1) = o(1),
\end{aligned}$$

by (25), (15) and (17).

We now prove (20). Wald's identity gives

$$\begin{aligned}
\mathsf{E}_\theta(S_T - b)^m &= \mathsf{E}_0\left[(S_T - b)^m \exp\{\theta S_T - \psi(\theta)T\}\right] \\
&= \mathsf{E}_0\left[(S_T - b)^m \exp\left\{\theta b + \theta(S_T - b) - \frac{T}{b^2}\frac{\psi(\theta)}{\theta^2/2}\frac{(\theta b)^2}{2}\right\}\right].
\end{aligned}$$

Hence, we get

(26)

$$\lim \mathsf{E}_\theta(S_T - b)^m = \lim \mathsf{E}_0\left[(S_T - b)^m \exp\left\{\theta b + \theta(S_T - b) - \frac{T}{b^2}\frac{\psi(\theta)}{\theta^2/2}\frac{(\theta b)^2}{2}\right\}\right].$$

Assuming for the moment that the expression $(S_T - b)^m \exp\{\theta(S_T - b)\}$ is uniformly integrable under F_0, we can pass the limit on the right-hand side of (26) inside the expectation. Let $\tau_B(1)$ denote the first time a standard Brownian motion process hits 1. Under P_0, (16) implies $b^{-2}T \Rightarrow \tau_B(1)$. In addition, we have $\theta b \to \zeta$; $\theta(S_T - b) \Rightarrow 0$ because $\theta \to 0$ and $S_T - b$ has a finite weak limit; and $2\psi(\theta)/\theta^2 \to 1$ by (4). These facts, plus (19), yield

$$\begin{aligned}
\lim \mathsf{E}_\theta(S_T - b)^m &= \mathsf{E}_0(R_\infty - \xi)^m \cdot \mathsf{E}_0\left[\exp\left\{\zeta - \frac{\zeta^2}{2}\tau_B(1)\right\}\right] \\
&= \mathsf{E}_0(R_\infty - \xi)^m,
\end{aligned}$$

where the second equality follows from Wald's identity for Brownian motion (as in Siegmund 1985, Proposition 3.2).

The uniform integrability of $(S_T - b)^m \exp\{\theta(S_T - b)\}$ used above is verified as follows. We let $\xi_+ = \sup_n \xi_n$, $\xi_- = \inf_n \xi_n$, and define the stopping time $T_- = \inf\{n \ge 1: S_n > b - \xi_-\}$. Observe that $S_T \le S_{T_-}$. Now we bound $(S_T - b)^m \exp\{\theta(S_T - b)\}$ by

$$C \cdot \exp\{(\theta + \varepsilon)(S_T - b)\}\mathbf{1}_{\{S_T > b\}} + (\xi_+)^m \mathbf{1}_{\{S_T \le b\}}, \tag{27}$$

for some $C > 0$ and some $\varepsilon > 0$. On the event $\{S_T > b\}$ the bound is clear; on the event $\{S_T \le b\}$ we use the fact that $|S_T - b| \le \xi_+$. Now (27) is bounded by

$$C \cdot \exp\{(\theta + \epsilon)(R_{b-\xi_-} - \xi_-)\} + (\xi_+)^m,$$

where $R_{b-\xi_-} = S_{T_-} - (b - \xi_-)$ is the excess for an ordinary random walk over level $b - \xi_-$. This bounding sequence is uniformly integrable by Lemma XII.6.4 of Asmussen (1987). The uniform integrability required above now follows by the dominated convergence theorem. □

3. The random walk at perturbed crossings. From now on, the perturbing terms $\{\xi_n\}$ are numbers satisfying (12), as they do in (11). The stopping times T and τ are as in (13) and (21). In this section, we prove two lemmas on S_T, the location of the random walk when the perturbed walk crosses a boundary.

LEMMA 2. *If F_0 is nonlattice, then for all sufficiently small $\theta^* > 0$ there exists $\alpha > 0$ such that*

$$\sup_{\theta_1 \in [0,\, \theta^*]} \left| \mathsf{E}_{\theta_1}[S_T - S_\tau] \right| = O(e^{-\alpha b})$$

as $b \to \infty$.

PROOF. We start from the inequality

$$\left| \mathsf{E}_{\theta_1}[S_T - S_\tau] \right| = \left| \mathsf{E}_{\theta_1}[(S_T - S_\tau); T \neq \tau] \right| \leq \sqrt{P_{\theta_1}\{T \wedge \tau < n^*\}}\, \sqrt{\mathsf{E}_{\theta_1}(S_T - S_\tau)^2}, \tag{28}$$

and bound each of the factors on the right.

We claim that each of the probabilities $P_{\theta_1}\{\tau < n^*\}$ and $P_{\theta_1}\{T < n^*\}$ is $O(e^{-\alpha_1 b})$, as $b \to \infty$, for some $\alpha_1 > 0$, uniformly in all sufficiently small θ_1. To see this, choose $\theta^* > 0$ and $\alpha^1 > 0$ so that $\psi(\theta^* + \alpha^1) < \infty$. Then for all $\theta_1 \in [0, \theta^*]$,

$$\begin{aligned} P_{\theta_1}\{\tau < n^*\} &\leq P_{\theta_1}\Big\{ \max_{1 \leq n \leq n^*} S_n > b - \xi \Big\} \\ &\leq \sum_{n=1}^{n^*} P_{\theta_1}\{S_n > b - \xi\} \\ &\leq e^{-\alpha_1(b-\xi)} \sum_{n=1}^{n^*} \mathsf{E}_{\theta_1} e^{\alpha_1 S_n} \\ &= e^{-\alpha_1(b-\xi)} \sum_{n=1}^{n^*} e^{n[\psi(\theta_1+\alpha)-\psi(\theta_1)]} \\ &\leq e^{-\alpha_1(b-\xi)} \sum_{n=1}^{n^*} e^{n\psi(\theta^*+\alpha_1)} \triangleq Ce^{-\alpha_1 b}. \end{aligned} \tag{29}$$

The argument for $P_{\theta_1}\{T < n^*\}$ is similar.

Now set $T' = \inf\{n \geq 1\colon S_n > b - \xi_*\}$, with $\xi_* = \min_n \xi_n$, and observe that $S_T \leq S_{T'}$ and $S_T \leq S_{T'}$. Applying (29) to (28) and twice using the inequality $(x + y)^2 \leq 2(x^2 + y^2)$, we find that

$$\begin{aligned} \sup_{\theta_1 \in [0,\, \theta^*]} \left| \mathsf{E}_{\theta_1}[S_T - S_\tau] \right| &\leq \sqrt{Ce^{-\alpha_1 b}} \sup_{\theta_1 \in [0,\, \theta^*]} \sqrt{2\left(\mathsf{E}_{\theta_1} S_T^2 + \mathsf{E}_{\theta_1} S_\tau^2\right)} \\ &\leq \sqrt{Ce^{-\alpha_1 b}} \sup_{\theta_1 \in [0,\, \theta^*]} \sqrt{4\mathsf{E}_{\theta_1} S_{T'}^2} \\ &\leq O(e^{-\alpha_1 b/2}) \sup_{\theta_1 \in [0,\, \theta^*]} \sqrt{8\left[(b - \xi_*)^2 + \mathsf{E}_{\theta_1} R_{b-\xi_*}^2\right]}. \end{aligned}$$

From Theorem 3 of Lorden (1970), we get

$$\mathsf{E}_{\theta_1} R_{b-\xi_*}^2 \le \frac{4}{3} \frac{\mathsf{E}_{\theta_1} S_{\tau_+}^3}{\mathsf{E}_{\theta_1} S_{\tau_+}}, \quad \text{for all } \theta_1 \ge 0.$$

From Siegmund (1979, p. 706), we know that the moments of S_{τ_+} are continuous in $\theta_1 \in [0, \theta^*]$ for sufficiently small θ^*, and $\mathsf{E}_{\theta_1} S_{\tau_+}$ is bounded away from 0. Thus,

$$\sup_{\theta_1 \in [0,\, \theta^*]} \sup_{b>0} \mathsf{E}_{\theta_1} R_{b-\xi_*}^2 < \infty,$$

and we conclude that $\sup_{\theta_1 \in [0,\, \theta^*]} |\mathsf{E}_{\theta_1}[S_T - S_\tau]|$ is $O(e^{-\alpha_1 b/2}) \cdot \sqrt{O(b^2) + O(1)}$, which is $O(e^{-\alpha b})$ for $\alpha = \alpha_1/2$. □

LEMMA 3. *If F_0 is strongly nonlattice and if $\theta_1 \downarrow 0$, $b \to \infty$ and $\theta_1 b \to$ constant, then*

$$\mathsf{E}_{\theta_1}[S_T - b] = \beta - \xi + \theta_1(\kappa - \beta^2) + o(\theta_1). \tag{30}$$

PROOF. Starting from the representation $\mathsf{E}_{\theta_1} R_\infty = \mathsf{E}_{\theta_1} S_{\tau_+}^2 / (2\mathsf{E}_{\theta_1} S_{\tau_+})$ and expanding both numerator and denominator according to Lemma 2 of Siegmund (1979), we arrive at

$$\mathsf{E}_{\theta_1} R_\infty = \beta + \theta_1(\kappa - \beta^2) + o(\theta_1). \tag{31}$$

Corollary 2.3 of Chang (1992) shows that for strongly nonlattice F_0,

$$\sup_{\theta_1 \in [0,\, \theta^*]} |\mathsf{E}_{\theta_1} R_b - \mathsf{E}_{\theta_1} R_\infty| = O(e^{-\alpha b}),$$

for some $\alpha > 0$. But then (31) holds with R_∞ replaced by $R_{b-\xi}$; more precisely,

$$\mathsf{E}_{\theta_1}[S_\tau - (b - \xi)] = \beta + \theta_1(\kappa - \beta^2) + o(\theta_1), \tag{32}$$

as $\theta_1 \downarrow 0$, $b \to \infty$ and $\theta_1 b \to$ constant. Equation (30) now follows from Lemma 2. □

4. Analysis of the approximations. We first prove part (ii) of Theorem 1, then parts (i), (iii), and (iv) and the corollary.

4.1. Stockout probability. For Theorem 1(ii), we write

$$P_{\theta_0}\{Y^1 > b\} = P_{\theta_0}\{T < \infty\} = \mathsf{E}_{\theta_1}[e^{-\gamma S_T}],$$

where the first equality follows from (13), and the second is Wald's identity. By Taylor expansion,

$$\begin{aligned} P_{\theta_0}\{Y^1 > b\} &= e^{-\gamma b} \mathsf{E}_{\theta_1}\left[1 - \gamma(S_T - b) + \frac{\gamma^2}{2}(S_T - b)^2 + O(\gamma^3)\right] \\ &= e^{-\gamma b}\left\{1 - \gamma \mathsf{E}_{\theta_1}[S_T - b] + \frac{\gamma^2}{2}\mathsf{E}_{\theta_1}(S_T - b)^2 + O(\gamma^3)\right\}. \end{aligned} \tag{33}$$

That the term $\mathsf{E}_{\theta_1}[O(\gamma^3)]$ is $O(\gamma^3)$ follows from the inequalities $1 - x + (x^2/2) - (x^3/6) \le e^{-x} \le 1 - x + x^2/2$ and the convergence of $\mathsf{E}_{\theta_1}(S_T - b)^3$ to a finite limit, as

ensured by (20). If F_0 is strongly nonlattice, substitute (30) and (20) with $m = 2$ into (33), recalling that $\beta = \mathsf{E}_0 R_\infty$, $\kappa = \mathsf{E}_0 R_\infty^2$, to get

$$\begin{aligned} P_{\theta_0}\{Y^1 > b\} &= e^{-\gamma b}\Big\{1 - \gamma\Big(\beta - \xi + \frac{\gamma}{2}(\kappa - \beta^2) + o(\gamma)\Big) \\ &\qquad + \frac{\gamma^2}{2}(\kappa - 2\xi\beta + \xi^2 + o(1)) + O(\gamma^3)\Big\} \\ &= e^{-\gamma(b+\beta-\xi)} + o(\gamma^2). \end{aligned}$$

For merely nonlattice F_0, (30) need not hold, but we still have

$$\mathsf{E}_{\theta_1}[S_T - b] = \beta - \xi + o(1),$$

according to (20). The approximation therefore becomes

$$\begin{aligned} P_{\theta_0}\{Y^1 > b\} &= e^{-\gamma b}\{1 - \gamma(\beta - \xi + o(1)) + o(\gamma)\} \\ &= e^{-\gamma(b+\beta-\xi)} + o(\gamma). \end{aligned}$$

4.2. Mean shortfall. We prove Theorem 1(i) by establishing the equivalent fact that

$$\mathsf{E}_{\theta_0} Y^1 = \gamma^{-1} - \beta + \xi + O(\gamma).$$

Suppose the maximum of the random walk S_n is attained at τ^* and that of the perturbed random walk Z_n at T^*. Let $W = \max_{n \geq 0} S_n = S_{\tau^*}$. As a consequence of (12), the maxima are attained simultaneously if both are attained after $n^* - 1$. Therefore, we have

$$\begin{aligned} (34)\quad \mathsf{E}_{\theta_0}[Y^1 - W - \xi] &= \mathsf{E}_{\theta_0}[S_{T^*} + \xi_{T^*} - S_{\tau^*} - \xi] \\ &\leq \mathsf{E}_{\theta_0}[(S_{T^*} + \xi_{T^*} - S_{\tau^*} - \xi); \tau^* < n^*] \\ &\quad + \mathsf{E}_{\theta_0}[(S_{T^*} + \xi_{T^*} - S_{\tau^*} - \xi); T^* < n^*] \\ &\leq \Big(\max_{1 \leq n \leq n^*} \xi_n - \xi\Big)\big(P_{\theta_0}\{\tau^* < n^*\} + P_{\theta_0}\{T^* < n^*\}\big). \end{aligned}$$

The equality uses $Y^1 \overset{\mathrm{d}}{=} S_{T^*} + \xi_{T^*}$, and the second inequality uses $S_{T^*} \leq S_{\tau^*}$.

Below we argue that $P_{\theta_0}\{\tau^* < n^*\}$ and $P_{\theta_0}\{T^* < n^*\}$ are $O(\gamma)$. Once this is established, (6) applied to $\mathsf{E}_{\theta_0} W$ proves part (i) via (34). It suffices to verify that $P_{\theta_0}\{T^* < n^*\}$ is $O(\gamma)$, because the claim for $P_{\theta_0}\{\tau^* < n^*\}$ is a special case ($\xi_n \equiv 0$).

Define strong ascending ladder epochs $\{T_+^{(k)}, k \geq 0\}$, ($T_+^{(0)} \equiv 0$) for Z_n just as they are defined for S_n (e.g., Asmussen 1987, p. 167). Then the perturbed walk attains its maximum at

$$T^* = \sup\{T_+^{(k)} : T_+^{(k)} < \infty, k \geq 0\}.$$

Let $\xi^* = \max_n \xi_n$ and $\xi_* = \min_n \xi_n$. If we define

$$t^{(1)} = \inf\{n \geq 1 : S_n + \xi_* > 0\}$$

then $t^{(1)} \geq T_+^{(1)}$, a.s., and therefore

$$P_{\theta_0}\{T_+^{(1)} < \infty\} \geq P_{\theta_0}\{t^{(1)} < \infty\} = \mathsf{E}_{\theta_1}[e^{-\gamma S_{t^{(1)}}}] \tag{35}$$

$$= \mathsf{E}_{\theta_1}[1 - \gamma S_{t^{(1)}} + o(\gamma)] = 1 + O(\gamma),$$

because $\lim_{\theta_1 \to 0} \mathsf{E}_{\theta_1} S_{t^{(1)}} = \mathsf{E}_0 S_{t^{(1)}} < \infty$. For $k \geq 2$, set

$$t^{(k)} = \inf\{n > T_+^{(k-1)} : S_n + \xi_* > S_{T_+^{(k-1)}} + \xi^*\}.$$

Observe that $t^{(k)} \geq T_+^{(k)}$, a.s., and

$$t^{(k)} \stackrel{\mathrm{d}}{=} t' \stackrel{\Delta}{=} \inf\{n > 0 : S_n > \xi^* - \xi_*\},$$

for all $k \geq 2$. Obviously,

$$P_{\theta_0}\{T_+^{(k)} < \infty\} \geq P_{\theta_0}\{t^{(k)} < \infty\} = P_{\theta_0}\{t' < \infty\} = 1 + O(\gamma). \tag{36}$$

Finally, by defining $N = \sup\{k \geq 0 : T_+^{(k)} < \infty\}$ we conclude that

$$P_{\theta_0}\{T^* < n^*\} \leq P_{\theta_0}\{N < n^*\}$$

$$= \sum_{k=0}^{n^*-1} P_{\theta_0}\{N = k\}$$

$$= \sum_{k=0}^{n^*-1} P_{\theta_0}\{T_+^{(1)} < \infty\} \cdots P_{\theta_0}\{T_+^{(k)} < \infty\} P_{\theta_0}\{T_+^{(k+1)} = \infty\}$$

$$= O(\gamma),$$

the last equality following from (35)–(36). □

4.3. Average backlog. Suppose F_0 is strongly nonlattice. In a single-node system we have $\xi_n \equiv 0$, so $Y^1 \stackrel{\mathrm{d}}{=} W = \max_{n \geq 0} S_n$, and the approximation for the average backlog becomes

$$\mathsf{E}_{\theta_0}(W - b)^+ = \frac{1}{\gamma} e^{-\gamma(b+\beta)} + o(\gamma), \tag{37}$$

where, as before, $W = \max_{n \geq 0} S_n$. We first prove (37), then use this approximation to prove part (iii) of Theorem 1.

PROOF OF (37). Using τ' defined in (14) and the strong Markov property, we write the average backlog as

$$\mathsf{E}_{\theta_0}(W - b)^+ = P_{\theta_0}\{W > b\} \mathsf{E}_{\theta_0}[W - b | W > b]$$

$$= P_{\theta_0}\{W > b\} \left(\mathsf{E}_{\theta_0} W + \mathsf{E}_{\theta_0}[S_{\tau'} - b | \tau' < \infty]\right).$$

The probability $P_{\theta_0}\{W > b\}$ can be approximated by (7) and $\mathsf{E}_{\theta_0}W$ by (6), whereas

$$
\begin{aligned}
(38)\qquad & P_{\theta_0}\{W > b\}\mathsf{E}_{\theta_0}[S_{\tau'} - b | \tau' < \infty] \\
&= \mathsf{E}_{\theta_0}[S_{\tau'} - b] \\
&= \mathsf{E}_{\theta_1}[(S_{\tau'} - b)e^{-\gamma S_{\tau'}}] \\
&= e^{-\gamma b}\Big\{\mathsf{E}_{\theta_1}[S_{\tau'} - b] - \gamma \mathsf{E}_{\theta_1}(S_{\tau'} - b)^2 + o(\gamma)\Big\} \\
&= e^{-\gamma b}\Big\{\Big(\beta + \frac{\gamma}{2}(\kappa - \beta^2) + o(\gamma)\Big) - \gamma(\kappa + o(1)) + o(\gamma)\Big\} \\
&= e^{-\gamma b}\Big\{\beta - \frac{\gamma}{2}(\kappa + \beta^2) + o(\gamma)\Big\}.
\end{aligned}
$$

Combining the approximation gives (37).

We now prove the general case of Theorem 1(iii). Suppose that the following hold as $b \to \infty$, $\theta_0 \uparrow 0$ and $b\theta_0 \to$ constant:

(a) $\mathsf{E}_{\theta_0}([W + \xi^* - b]^+)^2 = O(\gamma^{-2})$,

(b) $\mathsf{E}_{\theta_0}([W + \xi - b]^+)^2 = O(\gamma^{-2})$,

(c) $P_{\theta_0}\{\tau < n^*\} = O(e^{-\alpha_1 b})$, $\alpha_1 > 0$, and

(d) $P_{\theta_0}\{T < n^*\} = O(e^{-\alpha_1 b})$, $\alpha_1 > 0$.

With T^* the time Z_n achieves its maximum, write $\mathsf{E}_{\theta_0}(Y^1 - b)^+$ as $\mathsf{E}_{\theta_0}(Z_{T^*} - b)^+$ and decompose the latter as

$$
(39)\qquad \mathsf{E}_{\theta_0}\big[(Z_{T^*} - b)^+; T \wedge \tau < n^*\big] + \mathsf{E}_{\theta_0}\big[(Z_{T^*} - b)^+; T \geq n^*, \tau \geq n^*\big].
$$

For the first term in (39) we have

$$
\begin{aligned}
(40)\qquad & \mathsf{E}_{\theta_0}\big[(Z_{T^*} - b)^+; T \wedge \tau < n^*\big] \\
&\leq \mathsf{E}_{\theta_0}\big[(W + \xi^* - b)^+; T \wedge \tau < n^*\big] \\
&\leq \sqrt{\mathsf{E}_{\theta_0}([W + \xi^* - b]^+)^2}\,\sqrt{P_{\theta_0}\{T \wedge \tau < n^*\}} = O(e^{-\alpha b}),
\end{aligned}
$$

by (a), (c) and (d). For the second term, we have

$$
\begin{aligned}
(41)\qquad & \mathsf{E}_{\theta_0}\big[(Z_{T^*} - b)^+; T \geq n^*, \tau \geq n^*\big] \\
&= \mathsf{E}_{\theta_0}\big[(W - b + \xi)^+; T \geq n^*, \tau \geq n^*\big] \\
&= \mathsf{E}_{\theta_0}(W - b + \xi)^+ - \mathsf{E}_{\theta_0}\big[(W - b + \xi)^+; T \wedge \tau < n^*\big].
\end{aligned}
$$

It follows from (b)–(d) that $\mathsf{E}_{\theta_0}[(W - b + \xi)^+; T \wedge \tau < n^*]$ is $O(e^{-\alpha b})$. Theorem 1(iii) now follows from (39)–(41) and (37).

It remains to verify (a)–(d) above. Define $T'' = \inf\{n \geq 1\colon S_n > b - \xi^*\}$. By the strong Markov property,

$$\mathsf{E}_{\theta_0}\big([W + \xi^* - b]^+\big)^2 = \mathsf{E}_{\theta_0}\big[(W + \xi^* - b)^2; W > b - \xi^*\big] \tag{42}$$

$$= \mathsf{E}_{\theta_0}\big[(W' + R_{b-\xi^*})^2; T'' < \infty\big],$$

where W' has the same distribution as W and is independent of $(T'', R_{b-\xi^*})$. The expectation in (42) is no larger than

$$2\mathsf{E}_{\theta_0}W^2 \cdot P_{\theta_0}\{T'' < \infty\} + 2\mathsf{E}_{\theta_0}\big[R^2_{b-\xi^*}; T'' < \infty\big].$$

Now $\mathsf{E}_{\theta_0}W^2$ is indeed $O(\gamma^{-2})$ according to equation (10) in Siegmund (1979); $P_{\theta_0}\{T'' < \infty\}$ is $e^{-\gamma(b-\xi^*+\beta)} + o(\gamma^2)$ according to (7); and

$$\mathsf{E}_{\theta_0}\big[R^2_{b-\xi^*}; T'' < \infty\big] = e^{-\gamma(b-\xi^*)}\mathsf{E}_{\theta_1}\big[R^2_{b-\xi^*}e^{-\gamma R_{b-\xi^*}}\big]$$

$$= e^{-\gamma(b-\xi^*)}\mathsf{E}_{\theta_1}\big[R^2_{b-\xi^*} - \gamma R^3_{b-\xi^*} + o(\gamma)\big]$$

$$= e^{-\gamma(b-\xi^*)}[\kappa + o(1)].$$

Combining these approximations proves claim (a). Claim (b) follows automatically from (a). An argument similar to that leading to (29) proves (c) and (d).

If we assume merely that F_0 is nonlattice, then the approximation for the average backlog in a single-stage system becomes

$$\mathsf{E}_{\theta_0}(W - b)^+ = \frac{1}{\gamma}e^{-\gamma(b+\beta)} + o(1).$$

The lower-order error is a consequence of the fact that, in the argument used for (37), $P_{\theta_0}\{W > b\}$ is now replaced with $e^{-\gamma(b+\beta)} + o(\gamma)$, and (38) with $e^{-\gamma b}[\beta + o(1)]$. Appropriate modification of the proof above leads to an error of $o(1)$ in the multistage approximation as well.

4.4. Unfilled demand. The unfilled demand defined in §1 can alternatively be written as

$$u(b) = \mathsf{E}_{\theta_0}\big[(Y^1 + D - b)^+ - (Y - b)^+\big] \tag{43}$$

$$= \mathsf{E}_{\theta_0}\int_0^D P_{\theta_0}(Y^1 > b - y)\,dy.$$

The fact that $\mathsf{E}_{\theta_0}[\exp(\gamma D)] = \exp(\gamma c^*)$ implies that

$$\frac{1}{\gamma}e^{-\gamma(b+\beta-\xi)}(e^{\gamma c^*} - 1) = \mathsf{E}_{\theta_0}\int_0^D e^{-\gamma(b-y+\beta-\xi)}\,dy. \tag{44}$$

We therefore need to show that

$$B \overset{\Delta}{=} \left|\mathsf{E}_{\theta_0}\int_0^D P_{\theta_0}(Y^1 > b - y) - e^{-\gamma(b-y+\beta-\xi)}\,dy\right|$$

is $o(\gamma^{2-\epsilon})$ for strongly nonlattice F_0, and $o(\gamma^{1-\epsilon})$ for nonlattice F_0.

For any $0 < \epsilon < 1$, define

$$B_1 = \left| \mathsf{E}_{\theta_0}\left[\int_0^D P_{\theta_0}(Y^1 > b - y) - e^{-\gamma(b-y+\beta-\xi)}\, dy;\ D > b^\epsilon \right] \right|$$

and

$$B_2 = \int_0^{b^\epsilon} \left| P_{\theta_0}(Y^1 > b - y) - e^{-\gamma(b-y+\beta-\xi)} \right| dy;$$

then $B \le B_1 + B_2$. Since $\gamma b \to$ constant, B_1 is bounded by a constant times

$$\mathsf{E}_{\theta_0}[De^{\gamma D};\ D > b^\epsilon] = \mathsf{E}_{\theta_0}[D;\ D > b^\epsilon] + \gamma \mathsf{E}_{\theta_0}[D^2;\ D > b^\epsilon]$$

$$+ \frac{\gamma^2}{2} \mathsf{E}_{\theta_0}[D^3;\ D > b^\epsilon] + o(\gamma^2).$$

For each $k = 1, 2, 3$, and sufficiently small $\alpha > 0$,

$$\mathsf{E}_{\theta_0}[D^k;\ D > b^\epsilon] = e^{-\theta_0 c^*} \mathsf{E}_0[D^k e^{\theta_0 D - \psi(\theta_0)};\ D > b^\epsilon]$$

$$\le e^{-\theta_0 c^*} \mathsf{E}_0[D^k;\ D > b^\epsilon]$$

$$\le e^{-\alpha b^\epsilon} e^{-\theta_0 c^*} \mathsf{E}_0[D^k e^{\alpha D}] = O(e^{-\alpha b^\epsilon}),$$

which is $O(e^{-\alpha'/\gamma^\epsilon})$ for some $\alpha' > 0$, and, in particular, is $o(\gamma^2)$.

For the other term we have

$$B_2 \le b^\epsilon \sup_{y \in [b - b^\epsilon,\, b]} \left| P_{\theta_0}(Y^1 > y) - e^{-\gamma(y+\beta-\xi)} \right|. \tag{45}$$

For any sequence $\{y_b\}$ with $y_b \in [b - b^\epsilon, b]$ we have $\lim y_b \theta_0 = \lim b\theta_0$, so Theorem 1(ii) implies $P_{\theta_0}(Y^1 > y_b) = \exp(-\gamma(y_b + \beta - \xi)) + o(\gamma^2)$. But then the supremum in (45) is also $o(\gamma^2)$, from which it follows that B_2 (hence also B) is $o(\gamma^{2-\epsilon})$. Exactly the same argument establishes an error of $o(\gamma^{1-\epsilon})$ in the nonlattice case.

In a single-node system, $[Y^1 + D - c^*]^+ \overset{\mathrm{d}}{=} Y^1$, so (43) can be rewritten as

$$u(b) = \int_0^{c^*} P_{\theta_0}(Y^1 > b - y)\, dy.$$

The argument applied to (45) shows that the convergence of the integrand to $\exp(-\gamma(b - y + \beta))$ is uniform over $[0, c^*]$, and thus that the error in the approximation to $u(b)$ is $o(\gamma^2)$. □

4.5. Safety stock. We now prove Corollary 1.

From Theorem 1(ii), we have

$$P_{\theta_0}\{Y^1 > s_\delta^1\} = e^{-\gamma(s_\delta^1 + \beta - \xi)} + o(\gamma^2)$$

$$= \delta + o(\gamma^2).$$

The claim for $1 - f(s_\delta^1)$ with the first expression given for s_δ^1 follows the same way using (3) and Theorem 1(iv). With the second expression given for s_δ^1 use (9) to get

$$1 - f(s_\delta^1) = \left(\frac{1}{\gamma c^*} + \frac{1}{2c^{*2}}\right) e^{-\gamma(s_\delta^1 + \beta - \xi)} \left(\gamma c^* + \frac{(\gamma c^*)^2}{2} + O(\gamma^3)\right) + o(\gamma)$$

$$= \left(1 + \frac{\gamma}{2c^*}\right) e^{-\gamma(s_\delta^1 + \beta - \xi)} \left(e^{\gamma c^*/2} + O(\gamma^2)\right) + o(\gamma)$$

$$= \delta\left(1 + \frac{\gamma}{2c^*}\right) e^{-\gamma/2c^*} + o(\gamma)$$

$$= \delta + o(\gamma). \quad \square$$

Acknowledgement. Research supported by NSF Grants ECS-9216490 and DMI-9457189.

References

Asmussen, S. (1984). Approximations for the probability of ruin within finite time. *Scand. Actuar. J.* 31–57.

___ (1987). *Applied Probability and Queues*, Wiley, Chichester, England.

___ (1992). Phase-type representations in random walk and queueing problems. *Ann. Probab.* **20** 772–789.

Chang, J. (1992). On moments of the first ladder height of random walks with small drift. *Ann. Appl. Probab.* **2** 714–738.

Chung, K. L. (1974). *A Course in Probability Theory*, Academic Press, New York.

Clark, A. J., H. Scarf (1960). Optimal policies for a multi-echelon inventory problem. *Management Sci.* **6** 475–490.

Federgruen, A., P. Zipkin (1986). An inventory model with limited production capacity and uncertain demands, I: The average cost criterion. *Math. Oper. Res.* **11** 193–207.

Glasserman, P. (1993). *Bounds and asymptotics for planning critical safety stocks*. Working Paper, Columbia University, New York, NY. *Oper. Res.* (to appear).

___, S. Tayur (1994). The stability of a capacitated, multi-echelon production-inventory system under a base-stock policy. *Oper. Res.* **42** 913–925.

___, ___ (1996). A simple approximation for a multistage capacitated production-inventory system. *Naval Res. Logist.* **43** 41–58.

Gut, A. (1992). First-passage times for perturbed random walks. *Sequential Anal.* **11** 149–179.

Hogan, M. L. (1986). Comment on 'corrected diffusion approximations in certain random walk problems.' *J. Appl. Probab.* **23** 89–96.

Liu, T.-W. (1995). *Analysis of a Capacitated Multistage Production-Inventory Inventory System under a Base-Stock Policy*. Ph.D. Dissertation, Graduate School of Business, Columbia University, New York, NY.

Lorden, G. (1970). On excess over the boundary. *Ann. Math. Statist.* **41** 520–527.

Schraner, E. (1995). *Optimal capacity allocation and inventory policies for capacitated multi-echelon systems*. Working Paper, Department of Industrial Engineering, Stanford University, Stanford, CA.

Siegmund, D. (1979). Corrected diffusion approximations in certain random walk problems. *Adv. Appl. Probab.* **11** 701–719.

___ (1985). *Sequential Analysis: Tests and Confidence Intervals*, Springer-Verlag, New York.

Speck, C. J., J. van der Wal (1991a). *The capacitated multi-echelon inventory system with serial structure*: 1. *The push ahead effect*, Memorandum COSOR 91-39, Eindhoven University of Technology, Eindhoven, The Netherlands.

___, ___ (1991b). *The capacitated multi-echelon inventory system with serial structure*: 2. *An average cost approximation method*, Memorandum COSOR 91-40, Eindhoven University of Technology, Eindhoven, The Netherlands.

Van Houtum, G. J., K. Inderfurth, W. H. M. Zijm (1995). *Materials Coordination in Stochastic Multiechelon Systems*, Working Paper LPOM-95-16, University of Twente, The Netherlands.

P. Glasserman: Columbia Business School, 403 Uris Hall, Columbia University, New York, New York 10027; e-mail: pglasser@research.gsb.columbia.edu

T.-W. Liu: Rutgers University, Newark, New Jersey 07102