requires that one understand the general distinction between a stochastic process and its distribution, and the specific distinction between standard Brownian motion and the Wiener measure.

2. Prove Proposition (3.2), which says that a continuous VF function has zero quadratic variation.
3. Calculate the variance of the sum on the left side of (3.4) and show that this vanishes as $n \to \infty$.
4. Let $X$ be the coordinate process on $C$ as in §A.3 and let $\nu(t,A,\omega)$ be the occupancy measure for $X$, defined by (3.4). Consider the particular point $\omega \in C$ defined by $\omega(t) = (1-t)^2$, $t \geq 0$. Fix a time $t > 1$ and describe $\nu(t,\cdot,\omega)$ in precise mathematical terms. Observe that this measure on $(R,\mathscr{B})$ is absolutely continuous (with respect to Lebesgue measure) but its density is not continuous. This substantiates a claim made in §3.
5. Prove (7.3) and (7.4). This is just a matter of verification, using the definitions of conditional expectation and martingale.
6. Let $X$ be a continuous adapted process on some filtered probability space $(\Omega,\mathbb{F},P)$. Define $V_\beta(t)$ in terms of $X$ via (5.8) and (5.10). The converse of (5.11) that was invoked in proving the change of measure theorem (7.9) is the following: If $V_\beta$ is a martingale for each $\beta \in R$, then $X$ is a $(\mu,\sigma)$ Brownian motion on $(\Omega,\mathbb{F},P)$. The problem is to prove this, specializing to the case $\mu = 0$ and $\sigma = 1$. As a first step, observe that $X$ is a (0,1) Brownian motion on $(\Omega,\mathbb{F},P)$ *if and only if*

$$(*) \qquad P\{X_{t+s} - X_t \leq x|\mathscr{F}_t\} = \Phi(s^{-1/2}x)$$

for $x \in \mathrm{R}$ and $s,t \geq 0$. Then show that $(*)$ is equivalent to

$$E\{\exp \beta(X_{t+s} - X_t)|\mathscr{F}_t\} = e^{\beta^2 s/2}.$$

## REFERENCES


1. P. Billingsley (1968), *Convergence of Probability Measures*, Wiley, New York.
2. L. Breiman (1968), *Probability*, Addison-Wesley, Reading, Mass.
3. K. L. Chung and R. J. Williams (1983), *Introduction to Stochastic Integration*, Birkhäuser, Boston.
4. D. Freedman (1971), *Brownian Motion and Diffusion*, Holden-Day, San Francisco.
5. R. S. Liptser and A. N. Shiryayev (1977), *Statistics of Random Processes*, Vol. I, Springer-Verlag, New York.
6. H. L. Royden (1968), *Real Analysis* (2nd ed.), Macmillan, New York.


CHAPTER 2

# Stochastic Models of Buffered Flow

Consider a firm that produces a single durable commodity on a make-to-stock basis. Production flows into a finished goods inventory, and demand that cannot be met from stock on hand is simply lost, with no adverse effect on future demand. The price of the output good is fixed, and demand is viewed as an exogenous source of uncertainty. Similarly, we consider plant, equipment, and work force size to be fixed for now, but there may be uncertainty about actual production quantities because of mechanical failures, worker absenteeism, and so forth. This firm and its market, portrayed schematically in Figure 1, constitute what we call a two-stage flow system. It consists of an input process (production), an output process (demand), and an intermediate buffer storage (the finished goods inventory) that serves to decouple input and output. Many mathematical models of such flow systems have been developed, with some aimed at particular areas of application and some quite abstract in character. For a sampling of these models see Arrow–Scarf–Karlin (1958), Moran (1959), Cox–Smith (1961), and Kleinrock (1976).

The abstract language of input processes, output processes, and storage buffers will be used hereafter, but the content of the buffer will be called inventory, and readers will find that all our examples involve production systems. In this chapter we develop a crude model of buffered flow, making no attempt to portray physical structure beyond that apparent in Figure 1. Actually, two models will be advanced, one with infinite buffer capacity and one with finite capacity. In each case, system flows are represented by continuous stochastic processes. Thus our models have little relevance to systems where individual inventory items are physically or economically


**Figure 1.** A two-stage flow system.

significant, but for discrete item systems with high-volume flow, the continuity assumption may be viewed as a convenient and harmless idealization.

## §1. A SIMPLE FLOW SYSTEM MODEL

Assume that the buffer in Figure 1 has infinite capacity. To model the system, we take as primitive a constant $X_0 \geq 0$ and two increasing, continuous stochastic processes $A = \{A_t, t \geq 0\}$ and $B = \{B_t, t \geq 0\}$ with $A_0 = B_0 = 0$. Interpret $X_0$ as the initial inventory level, $A_t$ as the cumulative input up to time $t$, and $B_t$ as the cumulative *potential* output up to time $t$. In other words, $B_t$ is the total output that can be realized over the time interval $[0,t]$ *if the buffer is never empty;* more generally, $B_t - B_s$ is the maximum possible output over the interval $(s,t]$. If emptiness does occur, then some of this potential output will be lost. We denote by $L_t$ the amount of potential output lost up to time $t$ because of such emptiness, so *actual* output over $[0,t]$ is $B_t - L_t$. Setting

(1) $$X_t \equiv X_0 + A_t - B_t\ ,$$

the inventory at time $t$ is then given by

(2) $$Z_t \equiv X_0 + A_t - (B_t - L_t) = X_t + L_t\ .$$

Most of our attention will focus on this *inventory process* $Z = \{Z_t, t \geq 0\}$. It remains to define the lost potential output process $L$ in terms of primitive model elements, and for that we simply assume (or require) that

(3) $L$ is increasing and continuous with $L_0 = 0$ and
(4) $L$ increases only when $Z = 0$.

Conditions (3) and (4) together say that output is (by assumption) sacrificed in the minimum amounts consistent with the physical restriction

(5) $$Z_t \geq 0 \quad \text{for all } t \geq 0\ .$$



In the next section it will be shown that conditions (2) to (5) uniquely determine $L$ and further imply the concise representation

(6) $$L_t = \sup_{0 \leq s \leq t} X_s^-\ .$$

Because $X$ is defined in terms of primitive elements by (1), this completes the precise mathematical specification of our two-stage flow system model with infinite buffer capacity.

A critical feature of this construction is that $L$ and $Z$ depend on $A$ and $B$ only through their difference, so one may view $X$ as the sole primitive element of our system model. Borrowing a term from the economic theory of production, we shall hereafter refer to $X$ as a *netput process*. This same term will be used later in other contexts, always to describe a net of potential input less potential output. The development above requires that $X$ have continuous sample paths, but thus far no probabilistic assumptions have been imposed. The emphasis in this chapter is on construction of sample paths rather than on probabilistic analysis.

## §2. THE ONE-SIDED REGULATOR

Let $C \equiv C[0,\infty)$ as in §A.2 Elements of $C$ will often be called paths or trajectories rather than functions, and the generic element of $C$ will be denoted by $x = (x_t, t \geq 0)$. We now define mappings $\psi, \phi : C \to C$ by setting

(1) $$\psi_t(x) \equiv \sup_{0 \leq s \leq t} x_s^- \qquad \text{for } t \geq 0$$

and

(2) $$\phi_t(x) \equiv x_t + \psi_t(x) \qquad \text{for } t \geq 0.$$

For purposes of discussion, fix $x \in C$ and let $l \equiv \psi(x)$ and $z \equiv \phi(x) = x + l$. We shall say that $z$ is obtained from $x$ by imposition of a *lower control barrier* at zero. The mapping $(\psi,\phi)$ will be called the *one-sided regulator* with lower barrier at zero. The effect of this path-to-path transformation is shown graphically in Figure 2, where the dotted line is $-l_t$. Note that $l = 0$ and hence $z = x$ up until the first time $t$ at which $x_t = 0$. Thereafter $z_t$ equals the amount by which $x_t$ exceeds the minimum value of $x$ over $[0,t]$.

**(3) Proposition.** Suppose $x \in C$ and $x_0 \geq 0$. Then $\psi(x)$ is the unique function $l$ such that

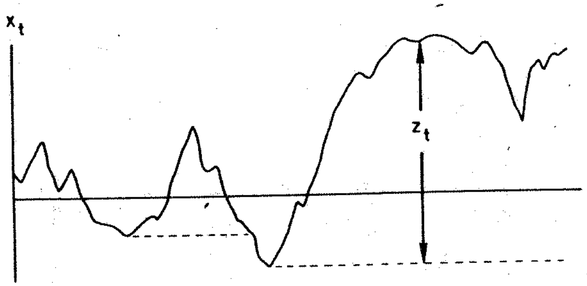


**Figure 2.** The one-sided regulator.

(4) $l$ is continuous and increasing with $l_0 = 0$,

(5) $z_t \equiv x_t + l_t \geq 0$ for all $t \geq 0$, and

(6) $l$ increases only when $z = 0$.

**(7) Remark.** Let $l$ be any function on $[0,\infty)$ satisfying (4) and (5) alone. It is easy to show that $l_t \geq \psi_t(x)$ for all $t \geq 0$. In this sense, the *least* solution of (4) and (5) alone is obtained by taking $l = \psi(x)$.

*Proof.* Fix $x \in C$ and set $l \equiv \psi(x)$ and $z \equiv x + l$. It is left as an exercise to show that this $l$ does in fact satisfy (4) to (6). To prove uniqueness, let $l^*$ be any other solution of (4) to (6) and set $z^* \equiv x + l^*$. Setting $y \equiv z^* - z = l^* - l$, we note that $y$ is a continuous VF function with $y_0 = 0$. Thus the Riemann–Stieltjes chain rule (B.4.1) gives

$$f(y_t) = f(0) + \int_0^t f'(y)\, dy \tag{8}$$

for any continuously differentiable $f: R \to R$. Taking $f(y) = y^2/2$, we see that (8) reduces to

$$\tfrac{1}{2}(z_t^* - z_t)^2 = \int_0^t (z^* - z)\, dl^* + \int_0^t (z - z^*)\, dl. \tag{9}$$

We know that $l^*$ increases only when $z^* = 0$, and $z \geq 0$, so the first term on the right side of (9) is $\leq 0$, and identical reasoning shows that the second term is $\leq 0$ as well. But because the left side is $\geq 0$, both sides must be zero. This shows that $z^* = z$ and hence $l^* = l$, and the proof is complete. □

Note that the property $l_0 = 0$ in (4) depends critically on the assumption that $x_0 \geq 0$. The following proposition shows that our one-sided regulator



has a sort of *memoryless property*. It will be used later to prove the strong Markov property of regulated Brownian motion.

**(10) Proposition.** Fix $x \in C$ and set $l \equiv \psi(x)$ and $z \equiv \phi(x) = x + l$. Fix $T > 0$ and define $x_t^* = z_T + (x_{T+t} - x_T)$, $l_t^* = l_{T+t} - l_T$, and $z_t^* = z_{T+t}$ for $t \geq 0$. Then $l^* = \psi(x^*)$ and $z^* = \phi(x^*)$.

Because the proof of (10) is just a matter of verification, it is left as an exercise. Pursuant to the observation (7), it is often helpful to think of $l_t$ as the cumulative amount of *control* exerted by an observer of the sample path $x$ up to time $t$. This observer must increase $l$ fast enough to keep $z \equiv x + l$ positive but wishes to exert as little control as possible subject to this constraint.

## §3. FINITE BUFFER CAPACITY

Consider again the two-stage flow system of §1, assuming now that the buffer has finite capacity $b$. Except as noted below, the assumptions and notation of §1 remain in force. In particular, the system netput process is defined by $X_t \equiv X_0 + A_t - B_t$, and $L_t$ denotes the amount of potential output lost up to time $t$ due to emptiness of the buffer. In the current context one must interpret $A$ as a *potential* input process; some of this potential input may be lost when the buffer is full. For reasons that will become clear in the next section, we denote by $U_t$ the total amount of potential input lost up to time $t$. Thus actual input up to time $t$ is $A_t - U_t$, and the inventory process $Z$ is given by

$$\begin{aligned} Z_t &= X_0 + (A_t - U_t) - (B_t - L_t) \\ &= X_t + L_t - U_t\,. \end{aligned} \tag{1}$$

Now how are $L$ and $U$ to be defined in terms of the primitive model elements? Assuming that $X_0 \in [0,b]$, it is more or less obvious from the development in §1 and §2 that $L$ and $U$ should be uniquely determined by the following properties:

(2) $L$ and $U$ are continuous and increasing with $L_0 = U_0 = 0$,

(3) $Z_t \equiv (X_t + L_t - U_t) \in [0,b]$ for all $t \geq 0$, and

(4) $L$ and $U$ increase only when $Z = 0$ and $Z = b$, respectively.

In the next section it will be shown that (2) to (4) do in fact determine $L$ and $U$ uniquely, although they cannot be expressed in neat formulas like (1.6).

Again a crucial point is that the processes of interest depend on primitive model elements only through the netput process $X$.

It is important to realize that a finite buffer may represent either a physical restriction on storage space or a policy restriction that shuts off input when buffer stock reaches a certain level. In the context of production systems, input is almost always controllable, and it is simply irrational to let inventory levels fluctuate without restriction. Thus the model described here is fundamentally more interesting than that developed in §1 and will be the focus of attention later.

## §4. THE TWO-SIDED REGULATOR

Fix $b > 0$ and let $C^*$ be the set of all functions $x \in C$ such that $x_0 \in [0,b]$. Given $x \in C^*$, we would like to find a pair of functions $(l,u)$ such that

(1) $l$ and $u$ are increasing and continuous with $l_0 = u_0 = 0$,

(2) $z_t \equiv (x_t + l_t - u_t) \in [0,b]$ for all $t \geq 0$, and

(3) $l$ and $u$ increase only when $z = 0$ and $z = b$, respectively.

Note that (3) associates $l$ and $u$ with the *lower* barrier at zero and *upper* barrier at $b$, respectively. If we consider $u$ to be given, then the requirements imposed on $l$ by (1) to (3) are those that define a lower control barrier at zero. That is, (1) to (3) and Proposition (2.3) together imply that

(4) $$l_t = \psi_t(x - u) \equiv \sup_{0 \leq s \leq t} (x_s - u_s)^- .$$

In exactly the same way, $u$ may be expressed in terms of $l$ via

(5) $$u_t = \psi_t(b - x - l) \equiv \sup_{0 \leq s \leq t} (b - x_s - l_s)^- .$$

It will now be proved that (4) and (5) together uniquely determine $l$ and $u$. The function $z$ defined by (2) may be pictured as in Figure 3, where the lower dotted line is $u_t - l_t$ and the upper dotted line is $b + u_t - l_t$. We shall henceforth say that $z$ is obtained from $x$ through imposition of a *lower control barrier at zero and an upper control barrier at b*.

**(6) Proposition.** For each $x \in C^*$, there is a unique pair of continuous functions $(l,u)$ satisfying (4) and (5), and this same pair uniquely satisfies (1) to (3).



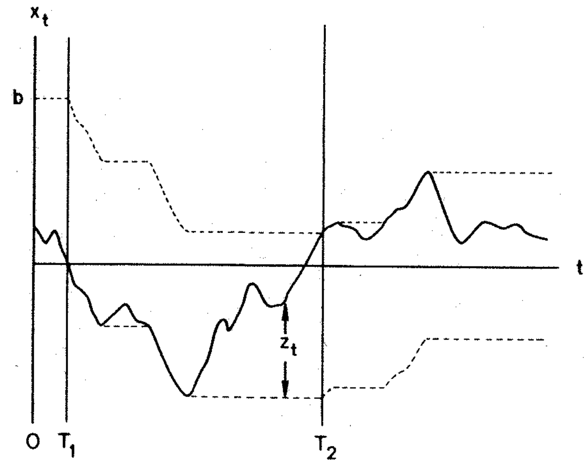


**Figure 3.** The two-sided regulator.

**(7) Definition.** We define mappings $f,g,h$: $C^* \to C$ by setting $f(x) \equiv l$, $g(x) \equiv u$, and $h(x) \equiv x + l - u$. Hereafter $(f,g,h)$ will be called the *two-sided regulator* with lower barrier at zero and upper barrier at $b$.

*Proof.* We first construct a solution of (4) and (5) by successive approximations. Beginning with the trial solution $l_t^0 \equiv u_t^0 \equiv 0$ $(t \geq 0)$, we set

(8) $$l_t^{n+1} \equiv \psi_t(x - u^n) \equiv \sup_{0 \leq s \leq t} (x_s - u_s^n)^-$$

and

(9) $$u_t^{n+1} \equiv \psi_t(b - x - l^n) \equiv \sup_{0 \leq s \leq t} (b - x_s - l_s^n)^-$$

for $n = 0, 1, \ldots$ and $t \geq 0$. Observe that $l_t^1 \geq l_t^0$ and $u_t^1 \geq u_t^0$ for all $t$, and hence (by induction) that $l_t^n$ and $u_t^n$ are increasing in $n$ for each fixed $t$. Thus we have

(10) $$l_t^n \uparrow l_t \quad \text{and} \quad u_t^n \uparrow u_t \text{ as } n \uparrow \infty .$$

Furthermore, it is easy to show that the convergence is achieved in a finite number of iterations for each fixed $t$, and the requisite number of iterations is an increasing function of $t$. For example, in Figure 3 we have $l_t = l_t^0$ and

$u_t = u_t^0$ if $0 \le t \le T_1$, $l_t = l_t^1$ and $u_t = u_t^1$ if $T_1 \le t \le T_2$, and so forth. (It is left as an exercise to show that $T_n \to \infty$, using the assumed continuity of $x$.) From this and (8) and (9) it follows that the limit functions $l$ and $u$ are finite valued, are continuous, and jointly satisfy (4) and (5).

To prove uniqueness, let $(l,u)$ and $(l^*,u^*)$ be two pairs of continuous functions satisfying (4) and (5), and let $z \equiv x + l - u$ and $z^* \equiv x + l^* - u^*$. From Proposition (2.3) it follows that $(l,u)$ and $(l^*,u^*)$ both satisfy (1) to (3) as well. Now let $y \equiv z^* - z = (l^* - l) - (u^* - u)$. Using the Riemann–Stieltjes chain rule as in the proof of Proposition (2.3), we find that

$$\tag{11} \tfrac{1}{2}(z_t^* - z_t)^2 = \int_0^t (z^* - z)\,dl + \int_0^t (z - z^*)\,dl^* + \int_0^t (z - z^*)\,du + \int_0^t (z^* - z)\,du^* .$$

Also as in the proof of Proposition (2.3), we use (1) to (3) to conclude that each term on the right side of (11) is $\le 0$, whereas the left side is $\ge 0$, and hence each side is zero. Thus $z^* = z$, from which it follows easily that $l^* = l$ and $u^* = u$ so that there is exactly one continuous pair $(l,u)$ satisfying (4) and (5). As we observed earlier, (1) to (3) and (4) and (5) are equivalent for continuous pairs $(l,u)$ by (2.3), and this proves the last statement of the proposition. □

**(12) Corollary.** For each fixed $t$, both $l_t \equiv f_t(x)$ and $u_t \equiv g_t(x)$ depend on $x$ only through $(x_s, 0 \le s \le t)$.

*Proof.* Immediate from the construction (8) to (10). □

**(13) Proposition.** Fix $x \in C$ and let $l \equiv f(x)$, $u \equiv g(x)$, and $z \equiv h(x)$ as above. Fix $T > 0$ and define $x_t^* \equiv z_T + (x_{T+t} - x_T)$, $l_t^* \equiv l_{T+t} - l_t$, $u_t^* \equiv u_{T+t} - u_T$, and $z_t^* = z_{T+t}$ for $t \ge 0$. Then $l^* = f(x^*)$, $u^* = g(x^*)$, and $z^* = h(x^*)$.

*Proof.* Starting with the fact that $x, l, u, z$ all satisfy (1) to (3), it is easy to verify that $x^*, l^*, u^*, z^*$ satisfy these same relations. The second uniqueness statement of (6) then establishes the desired proposition. □

## §5. MEASURING SYSTEM PERFORMANCE

In the design and operation of buffered flow systems, one is typically concerned with a tradeoff between system throughput characteristics and



the costs associated with inventory. Generally speaking, one can decrease the amount of lost potential input and output (which amounts to improving capacity utilization) by tolerating larger buffer stocks, but such stocks are costly in their own right.

To put the discussion on a concrete footing, consider again the single-product firm described at the beginning of this chapter. Recall that production flows into a finished goods inventory, and demand that cannot be met from stock on hand is simply lost with no adverse effect on future demand. Let $\pi$ denote the selling price (in dollars per unit of production) and let $B_t$ denote total demand over the time interval $[0,t]$. The latter notation is chosen for consistency with previous usage in §1 and §3.

Assuming plant and equipment are fixed, suppose that the firm must select at time zero a work force size, or equivalently a regular-time production capacity. For simplicity, assume that the work force size cannot be varied thereafter, the firm being obliged to pay workers their regular wages regardless of whether they are productively employed. Let $k$ be the capacity level selected, in units of production per unit time. The firm then incurs a labor cost of $wk$ dollars per unit time ever afterward, where $w > 0$ is a specified wage rate, even if it occasionally chooses to operate below capacity. For current purposes, overtime production is assumed to be impossible (see Problem 8). In addition to its labor costs, the firm incurs a materials cost of $m$ dollars per unit of *actual* production. Given the initial capacity decision (work force level), labor costs are fixed, and thus the *marginal cost of production is m dollars per unit.* A physical holding cost of $p$ dollars is incurred per unit time for each unit of production held in inventory. This includes such costs as insurance and security; it does *not* include the financial cost of holding inventory. (By financial cost we mean the opportunity loss on money tied up in inventory. More will be said on this subject shortly.)

It is assumed that the firm earns interest at rate $\lambda > 0$, compounded continuously, on funds that are not required for production operations. Continuous compounding means that one dollar invested at time zero returns $\exp(\lambda t)$ dollars of principal plus interest at time $t$. Thus a cost or revenue of one dollar at time $t$ is equivalent in value to a cost or revenue of $\exp(-\lambda t)$ dollars at time zero. Finally, we assume that the cumulative demand process $B$ satisfies

(1) $E(B_t) = at$ for all $t \ge 0$ $(a > 0)$ and

(2) $e^{-\lambda t} B_t \to 0$ almost surely as $t \to \infty$.

For one specific demand model that satisfies (1) and (2), we may suppose that the time axis can be divided into periods of unit length, that demand increments during successive periods form a sequence of independent and identically distributed random variables with mean $a$ and finite variance, and that demand arrives at a constant rate during each period. For this *linearized random walk* model of demand, property (1) is obvious and (2) follows from the strong law of large numbers. (The proof of this statement is left as an exercise.)

The firm must choose a capacity level $k$ at time zero and then at each time $t \geq 0$ select a production rate from the interval $[0,k]$. When a production rate below $k$ is selected, we shall say that *undertime* is being employed. For purposes of initial discussion, let us assume that management follows a *single-barrier policy* for production control after time zero. This means that production continues at the capacity rate $k$ until inventory hits some chosen level $b > 0$, and then undertime is employed in the minimum amounts necessary to keep inventory at or below level $b$. With this policy, our make-to-stock production system is a two-stage flow system with finite buffer capacity (see §3); the potential input process is $A_t \equiv kt$, and potential output is given by the demand process $B$. In the current context, $Z_t$ represents the finished goods inventory level at time $t$, $L_t$ is the cumulative demand lost up to time $t$, and $U_t$ is the cumulative undertime worked (potential production foregone) up to time $t$.

The firm's objective is to maximize the expected present value of sales revenues received minus operating expenses incurred over an infinite planning horizon, where discounting is continuous at interest rate $\lambda$. The actual production and sales volumes up to time $t$ are given by $kt - U_t$ and $B_t - L_t$, respectively; thus this amounts to maximization of

(3) $$V \equiv E\left[\pi \int_0^\infty e^{-\lambda t}(dB - dL) - wk \int_0^\infty e^{-\lambda t}\,dt - m \int_0^\infty e^{-\lambda t}(k\,dt - dU) - p \int_0^\infty e^{-\lambda t} Z_t\,dt\right],$$

where the integrals involving $dB$, $dL$, and $dU$ are defined path by path in the Riemann–Stieltjes sense (see Appendix B). The first term inside the expectation in (3) represents the present value of sales revenues, the second is the present value of labor costs, the third term is the present value of material costs (incremental production costs), and the last is the present value of inventory holding costs. It should be emphasized that the opportunity loss on capital tied up in inventory is fully accounted for by the discounting in (3); therefore $p$ should include only out-of-pocket expenses associated with holding inventory. To put it another way, no explicit financial cost of holding



inventory appears in (3) and including such a cost would be double counting. In a moment, however, we shall derive an equivalent measure of system performance in which a financial cost of inventory *does* appear. Readers who are not familiar with present value manipulations, and skeptical as to the appropriateness of (3) as a performance measure, may wish to consult §6.5. There is is shown that maximization of a discounted measure like $V$ is equivalent to maximizing the firm's expected total assets at a distant time of reckoning.

It will now be shown that maximization of $V$ is equivalent to minimization of another, somewhat simpler, performance measure. As a first step, consider the ideal situation where $B_t = at$ for all $t \geq 0$, meaning that demand arrives deterministically at constant rate $a$. We shall assume that

(4) $$\pi - w - m > 0\,,$$

for otherwise the system optimization problem would be uninteresting. (If $\pi - w - m \leq 0$, it is best to set $k = 0$ and go out of business.) With deterministic demand, one would, of course, choose $k = a$, meaning that units are produced precisely as demanded, labor and materials are paid for only as required for such production, and no inventory is held. The corresponding *ideal profit level* (in present value terms) would be

(5) $$I \equiv \int_0^\infty e^{-\lambda t}(\pi - w - m)a\,dt = \frac{(\pi - w - m)a}{\lambda}.$$

Now actual system performance under an arbitrary operating policy will be measured incrementally from this ideal. First, let

(6) $$\mu \equiv k - a,$$

(7) $$\delta \equiv \pi - m, \quad \text{and} \quad h \equiv p + m\lambda\,.$$

We call $\mu$ the *excess capacity*; it is the amount (possibly negative) by which chosen capacity exceeds the average demand rate. Interpret $\delta$ as a *contribution margin*; once the capacity level is fixed, each unit of sales contributes $\delta$ dollars to profit and the coverage of fixed costs. Finally, $h$ may be viewed as the *effective cost of holding inventory*; it consists of physical holding costs plus an opportunity loss rate of $\lambda$ times the marginal production cost $m$. It is assumed hereafter that $Z_0 = 0$.

**(8) Proposition.** $V = I - \Delta$, where

(9) $$\Delta \equiv E\left[\int_0^\infty e^{-\lambda t}(\delta\, dL + w\mu\, dt + hZ_t\, dt)\right].$$

**(10) Remark.** Because demand is exogenous, $I$ is an uncontrollable constant, and thus our original objective of maximizing $V$ is equivalent to minimizing $\Delta$.

*Proof.* From (1) and (2) it follows that

(11) $$E\left(\int_0^\infty e^{-\lambda t}\, dB\right) = E\left(\int_0^\infty \lambda e^{-\lambda t} B_t\, dt\right) = \int_0^\infty \lambda e^{-\lambda t} at\, dt = \int_0^\infty e^{-\lambda t} a\, dt.$$

The proof of (11), using Fubini's theorem and the Riemann–Stieltjes integration by parts theorem, is left as an exercise. Using (11), we can rewrite (5) as

(12) $$I = E\left[\int_0^\infty e^{-\lambda t}(\pi\, dB - wa\, dt - m\, dB)\right].$$

Now subtracting (3) from (12) we get

(13) $$I - V = E\left\{\int_0^\infty e^{-\lambda t}[\pi\, dL + w(k-a)\, dt + pZ_t\, dt + m(k\, dt - dU - dB)]\right\}.$$

With $Z_0 = 0$, we have $Z_t = (kt - U_t) - (B_t - L_t)$. Using this and integration by parts again, we find that

(14) $$\int_0^\infty e^{-\lambda t}(k\, dt - dU - dB) = \int_0^\infty e^{-\lambda t}(dZ - dL) = \int_0^\infty e^{-\lambda t}(\lambda Z_t\, dt - dL).$$

Substituting (14) into (13) and collecting similar terms, we have $I - V = \Delta$. □



Obviously $\Delta$ represents the amount by which management's plan falls short, in expected present value terms, of the ideal profit level $I$. The definition (9) expresses this shortfall as the sum of three effects. First, the contribution margin $\delta$ is lost on each unit of potential sales foregone. Second, we continuously incur a cost of $w$ dollars for each unit of capacity in excess of the average demand rate. Finally, for each unit of production held in inventory, we continuously incur an out-of-pocket cost $p$ plus an opportunity cost $\lambda m$. We emphasize again that $\Delta$ measures the *degradation of system performance from a deterministic ideal*. Thus the minimum achievable $\Delta$ value may be viewed as the *cost of stochastic variability*.

Our first objective here is to develop a quantitative theory of flow system performance. As a natural outgrowth of that descriptive objective, we also seek to prescribe means by which management can minimize or at least reduce performance degradation, such as investment in excess capacity (a design decision) and maintenance of buffer stock (a matter of operating policy).

In concluding this section, let us briefly consider a cost structure in which $\lambda \downarrow 0$ but $\delta$, $w$, and $h$ remain constant. Further suppose that

(15) $$\frac{1}{t}E(L_t) \to \alpha \text{ and } E(Z_t) \to \gamma \text{ as } t \to \infty.$$

Obviously $\alpha$ represents a long-run average *lost sales rate*, whereas $\gamma$ is the long-run average inventory level. Under mild additional assumptions, it is well known that $\lambda\Delta$ approaches the *long-run average cost rate*

(16) $$\rho \equiv \delta\alpha + w\mu + h\gamma$$

as $\lambda \downarrow 0$. Thus minimization of $\Delta$ is approximately equivalent to minimization of $\rho$ for small values of $\lambda$, and it is usually easier to calculate $\rho$ than the discounted performance measure $\Delta$.

## §6. BROWNIAN FLOW SYSTEMS

Suppose that, in the setting of §3, we directly model the netput process $X$ as a $(\mu,\sigma)$ Brownian motion. The inventory process $Z$, lost potential output $L$, and lost potential input $U$ are then defined by applying the two-sided regulator to $X$ exactly as before. In the obvious way, we call $Z$ a *regulated Brownian motion*, and the triple $(L,U,Z)$ will be referred to hereafter as a *Brownian flow system*. It will be seen later that all the performance measures discussed in §5, and a number of other interesting quantities, can be calculated explicitly for Brownian flow systems.

Although the Brownian system model is tractable, and therefore appealing, it is actually inconsistent with the model description given in §3; we have seen earlier that the sample paths of Brownian motion have infinite variation and thus it cannot represent the difference between a potential input process and a potential output process. Nonetheless, a netput process may be well approximated by Brownian motion under certain conditions. To understand these conditions, recall that Brownian motion is the *unique* stochastic process having stationary, independent increments and continuous sample paths; unbounded variation follows as a consequence of these primitive properties. Also note that the total variation of a netput process over any given interval equals the *sum* of potential input and potential output over that interval. If such a netput process is to be well approximated by Brownian motion, both potential input and potential output must be large for intervals of moderate length, but their difference (netput itself) must be moderate in value. We may express this state of affairs by saying that we have a system of *balanced high-volume flows*.

Pulling together several times, we conclude that Brownian motion may reasonably approximate the netput process for a system of *stationary, continuous, balanced high-volume flow, where netput increments during non-overlapping intervals are approximately independent*. Formal limit theorems that give this statement precise mathematical form, and thus serve to justify Brownian approximations, have been proved for various types of flow system models. The Brownian flow system will be studied extensively in future chapters, and readers should keep in mind its domain of applicability.

## PROBLEMS AND COMPLEMENTS

1. Prove Proposition (2.10), thus verifying the one-sided regulator's lack of memory.
2. Prove that $l \equiv \psi(x)$ satisfies (2.4) to (2.6).
3. Consider the three-stage flow system, or tandem buffer system, pictured in Figure 4. Each buffer has infinite capacity, and we denote by $X_k(0)$ the initial inventory in buffer $k$. Extending in an obvious way the model of §1, we take as primitive three increasing, continuous processes $A_k = \{A_k(t),\ t \geq 0\}$ such that $A_k(0) = 0$ ($k = 1,2,3$). Interpret $A_1$ as input to the first buffer, $A_2$ as potential transfer between the two buffers, and $A_3$ as potential output from the second buffer. Define a (continuous) vector netput process $X(t) = [X_1(t), X_2(t)]$ by setting



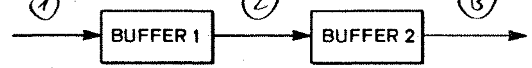


**Figure 4.** A three-stage flow system.

$$X_1(t) = X_1(0) + A_1(t) - A_2(t) \qquad \text{for } t \geq 0$$

and

$$X_2(t) = X_2(0) + A_2(t) - A_3(t) \qquad \text{for } t \geq 0 .$$

Let $L_2(t)$ denote the amount of the potential transfer $A_2(t)$ that is lost over $[0,t]$ because of emptiness of the first buffer, and define $L_3(t)$ in the obvious analogous fashion. Let $Z_k(t)$ denote the content of buffer $k$ at time $t$. Applying the analysis of §1 and §2 first to buffer 1 and then to buffer 2 in isolation, show that $L_2 = \psi(X_1)$, $Z_1 = \phi(X_1)$, $L_3 = \psi(X_2 - L_2)$, and $Z_2 = \phi(X_2 - L_2)$. Conclude that $L \equiv (L_2, L_3)$ and $Z \equiv (Z_1, Z_2)$ uniquely satisfy

(a) $L_2$ and $L_3$ are increasing and continuous with $L_2(0) = L_3(0) = 0$.
(b) $Z_1(t) = X_1(t) + L_2(t) \geq 0$ for all $t \geq 0$,
$Z_2(t) = X_2(t) - L_2(t) + L_3(t) \geq 0$ for all $t \geq 0$.
(c) $L_2$ and $L_3$ increase only when $Z_1 = 0$ and $Z_2 = 0$, respectively.

All of this describes the mapping by which $(L,Z)$ is obtained from $X$. (It is again important that $L$ and $Z$ depend on primitive model elements only through the netput process $X$.) Conditions (a) to (c) suggest the following interpretation or animation of that path-to-path transformation. An observer watches $X = (X_1, X_2)$ and may increase at will either component of a cumulative control process $L = (L_2, L_3)$. These actions determine $Z = (Z_1, Z_2)$ according to (b). The observer increases $L_2$ only as necessary to ensure that $Z_1 \geq 0$, so $L_2$ increases only when $Z_1 = 0$. Each such increase causes a positive displacement of $Z_1$ (or rather prevents a negative one) *and an equal negative displacement of* $Z_2$. Thus the effect of the observer's action at $Z_1 = 0$ is to drive $Z$ in the diagonal direction pictured in Figure 5. On the other hand, $L_3$ is increased at the boundary $Z_2 = 0$ so as to ensure $Z_2 \geq 0$, producing only the vertical displacement picture in Figure 5. Hereafter we shall say that $(L,Z)$ is obtained by applying a *multidimensional regulator* to $X$, the control region and *directions of control* being as illustrated in Figure 5. This problem is adapted from Harrison (1978).

4. A similar sort of multidimensional flow system is pictured in Figure 6. Here there are two input processes, each feeding its own infinite storage buffer. These inputs are then combined, exactly one unit of each input being required to produce one unit of system output. (The important

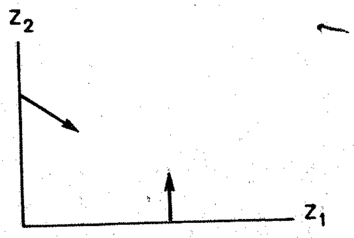


**Figure 5.** Directions of control for a three-stage flow system.

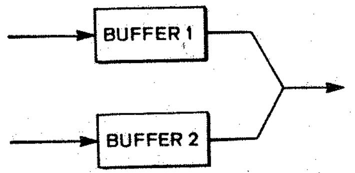


**Figure 6.** An assembly or blending operation.

point here is that inputs are combined in *fixed proportions*; the rest is just a matter of how units are defined.) This is the structure of an assembly operation, but again we treat the system flows as if they were continuous so that attention is effectively restricted to high-volume assembly systems. For another application, Figure 6 might be interpreted as a blending operation in which liquid or granulated ingredients are combined in fixed proportions to produce a similarly continuous output. To build a model, we again take as primitive initial inventory levels $X_1(0) \geq 0$ and $X_2(0) \geq 0$ plus three increasing, continuous processes $A_k = \{A_k(t),\ t \geq 0\}$ with $A_k(0) = 0$ $(k = 1,2,3)$. Interpret $A_1$ and $A_2$ as input to buffer 1 and buffer 2, respectively, and $A_3$ as potential output. Potential output is lost if *either* buffer is empty, and we denote by $L(t)$ the cumulative potential output lost up to time $t$ because of such emptiness. For purposes of determining $L$, the blending operation may be viewed as a two-stage flow system with initial inventory plus cumulative input given by

$$A^*(t) \equiv [X_1(0) + A_1(t)] \wedge [X_2(0) + A_2(t)] \,.$$

Let $Z_k(t)$ denote the inventory level in buffer $k$ at time $t$, and define a (continuous) vector netput process $X = [X_1(t), X_2(t)]$ by setting

$$X_1(t) \equiv X_1(0) + A_1(t) - A_3(t) \qquad \text{for } t \geq 0$$

and

$$X_2(t) \equiv X_2(0) + A_2(t) - A_3(t) \qquad \text{for } t \geq 0.$$



Applying the results of §1 and §2, write out explicit formulas for $L$ and $Z \equiv (Z_1, Z_2)$ in terms of $X$. (Again it is important that $L$ and $Z$ depend on primitive model elements only through the netput process $X$.) Conclude that $L$ and $Z$ together uniquely satisfy

(a) $L$ is continuous and increasing with $L(0) = 0$.
(b) $Z_1(t) = X_1(t) + L(t) \geq 0$ for all $t \geq 0$,
$Z_2(t) = X_2(t) + L(t) \geq 0$ for all $t \geq 0$.
(c) $L$ increases only when $Z_1 = 0$ or $Z_2 = 0$.

The mapping that carries $X$ into $(L,Z)$ may be pictured as in Figure 7. The inventory process $Z$ coincides with $X$ up until $X$ hits the boundary of the positive quadrant. At that point, $L$ increases, causing *equal positive displacements in both $Z_1$ and $Z_2$* as necessary to keep $Z_1 \geq 0$ and $Z_2 \geq 0$. Thus the effect of increases in $L$ at the boundary is to drive $Z$ in the diagonal direction shown in Figure 7, regardless of which boundary surface is struck. This problem is adapted from Harrison (1973).

**5.** Assuming for convenience that $x_0 = 0$, write out an explicit recursive expression for the times $T_1 < T_2 < \cdots$ identified in the proof of Proposition (4.6). Show that if $T_n \uparrow T < \infty$, then $x$ cannot be continuous at $T$; thus $T_n \to \infty$ as $n \to \infty$.

**6.** Consider again the three-stage flow system of Problem 3, assuming that buffers 1 and 2 now have *finite* capacities $b_1$ and $b_2$, respectively. In this case, potential input is lost when the first buffer is full, and potential transfer is lost when *either* the first buffer is empty *or* the second one is full. (We say that the transfer process is *starved* in the former case and *blocked* in the latter.) In addition to the notation established in Problem 3, let $L_1(t)$ denote total potential input lost up to time $t$. Argue that $L \equiv (L_1, L_2, L_3)$ and $Z \equiv (Z_1, Z_2, Z_3)$ should jointly satisfy

(a) $L_k$ is continuous and increasing with $L_k(0) = 0$ $(k = 1,2,3)$.
(b) $Z_1(t) \equiv X_1(t) + L_2(t) - L_1(t) \in [0,b_1]$ for all $t \geq 0$,
$Z_2(t) \equiv X_2(t) + L_3(t) - L_2(t) \in [0,b_2]$ for all $t \geq 0$.

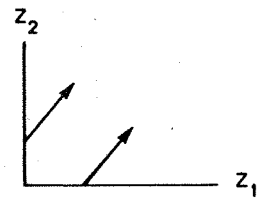


**Figure 7.** Directions of control for a blending operation.

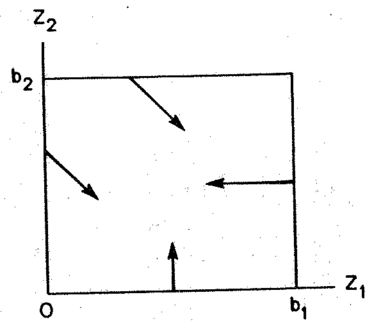


**Figure 8.** Directions of control for a three-stage flow system with finite buffers.

(c) $L_1$ increases only when $Z_1 = b_1$,
$L_2$ increases only when $Z_1 = 0$ or $Z_2 = b_2$, and
$L_3$ increases only when $Z_2 = 0$.

Explain the connection between (a) to (c) and Figure 8. Describe informally how one can use the results of Problem 3 and 4 to prove existence and uniqueness of a pair $(L,Z)$ satisfying (a) to (c). This problem is adapted from Wenocur (1982).

**7.** Show that the linearized random walk model of demand, described in §5, satisfies (5.1) and (5.2).

**8.** It was assumed in §5 that overtime production was impossible. Suppose instead that unlimited amounts of overtime production are available at a premium wage rate $w^* > w$, regardless of what work force level may be chosen at the beginning. To keep things simple, assume that overtime production is instantaneous. (One may also think in terms of buying finished goods at a premium price from some alternate supplier and then using these goods to satisfy demand.) Finally, assume that $\pi - w^* - m > 0$, so it is always better to use overtime production than to forego potential sales. The basic structure of this system is identical to that discussed in §5, but now $L_t$ is interpreted as cumulative overtime production up to time $t$. Show that maximizing the expected present value of total profit is equivalent to minimizing $\Delta$, where $\Delta$ is given by formula (5.9) with $w^*$ in place of $\delta$.

**9.** Prove the three equalities of (5.11), using Fubini's theorem (§A.5) and the Riemann–Stieltjes integration by parts theorem (§B.3).

## REFERENCES

1. K. J. Arrow, S. Karlin, and H. Scarf (1958), *Studies in the Mathematical Theory of Inventory and Production*, Stanford University Press, Stanford, Calif.



2. D. R. Cox and W. L. Smith (1961), *Queues*, Methuen, London.
3. J. M. Harrison (1973), Assembly-like Queues, *J. Appl. Prob.*, **10**, 354–367.
4. J. M. Harrison (1978), "The Diffusion Approximation for Tandem Queues in Heavy Traffic," *Adv. Appl. Prob.*, **10**, 886–905.
5. L. Kleinrock (1976), *Queuing Systems*, Vols. I and II, Wiley-Interscience, New York.
6. P. A. P. Moran (1959), *Theory of Storage*, Methuen, London.
7. M. Wenocur (1982), "A Production Network Model and Its Diffusion Limit," Ph.D. thesis, Statistics Department, Stanford University.