Manufacturing & Service Operations Management

Publication details, including instructions for authors and subscription information:
http://pubsonline.informs.org

Measuring Item Fill-Rate Performance in a Finite Horizon

Douglas J. Thomas,

M&SOM MANUFACTURING & SERVICE OPERATIONS MANAGEMENT
Vol. 7, No. 1, Winter 2005, pp. 74–80
ISSN 1523-4614 | EISSN 1526-5498 | 05 | 0701 | 0074

DOI 10.1287/msom.1040.0064

Measuring Item Fill-Rate Performance in a Finite Horizon

Douglas J. Thomas
Department of Supply Chain and Information Systems, Smeal College of Business Administration,
The Pennsylvania State University, University Park, Pennsylvania 16802, dthomas@psu.edu

The standard treatment of fill rate relies on stationary and serially independent demand over an infinite horizon. Even if demand is stationary, managers are held accountable for performance over a finite horizon. In a finite horizon, the fill rate is a random variable. Studying the distribution is relevant because a vendor may be subject to financial penalty if she fails to achieve her target fill rate over a specified finite period. It is known that for a zero lead time, base-stock model, the expected value of a finite-horizon fill rate exceeds the long-run fill rate. In this paper, I investigate the behavior of the distribution of the finite-horizon fill rate when a stationary base-stock policy is used to control inventory. For a vendor facing a finite-horizon, fill-rate-level contract and using a stationary stocking policy, I examine how the the length of the review horizon (i.e., monthly or quarterly), the demand distribution, and the cost of failing to meet the target affect the stocking decision.

Key words: inventory; base-stock policy; service-level constraint
History: Received: September 13, 2002; accepted: October 20, 2004. This paper was with the author 7 months for 4 revisions.

1. Introduction

The item fill rate, defined as the fraction of demand satisfied from stock, is a widely used measure of customer service for inventory systems. Methods presented in textbooks and used in commercial software for setting stock levels to achieve a target fill rate often assume that demand is stationary and serially independent for an infinite horizon, and replenishment orders arrive in the order in which they were placed. When these assumptions are satisfied, one can analyze a single-order replenishment cycle, and the fill rate converges to the quantity given by expected demand *satisfied* per cycle, divided by expected demand per cycle.

In a finite horizon, the fill rate is a random variable. The motivation to study this random variable comes from observing performance contracts in practice. An inventory manager may have a contract with a customer that a minimum fill rate will be realized over some finite review horizon, perhaps monthly or quarterly. A supplier may incur a penalty for not meeting this target, but may also receive a bonus for exceeding another (higher) target. The penalty for failing to meet a target might be a specific financial penalty imposed by a customer or simply one element on a supplier scorecard, with some implicit (but hard to quantify) cost of failure. Even when an external contract is not in place, a manager's performance will be reviewed internally over some finite horizon. This implies that a manager is—or at least should be—interested in the probability of meeting or exceeding a specified service level over a finite horizon rather than (or at least in addition to) the long-run performance.

Managers are of course aware that the achieved fill rate is often different from the level implied by the chosen stock level. Unfortunately, in many cases, explicit calculation of the distribution of fill rate, or the probability of achieving a certain target for a given stock level and horizon, is difficult as we will see below. As a result, some managers facing contractual service-level commitments use a traditional, infinite-horizon model, but "overshoot" to improve their chances of success. That is, if they need to exceed a 95% fill rate to avoid penalty, they might set the stock level based on a 97% long-run fill rate.

It should be noted that a variety of factors can contribute to the random behavior of the fill rate in a finite horizon. In addition to the finite-horizon effects explored here, the demand distributions (forecasts) may be incorrectly specified. Tyworth and O'Neill (1997) investigate the sensitivity of service performance by the shape of the lead time demand distribution in a continuous review system. Most stochastic inventory models assume that orders from different replenishment cycles do not cross. Robinson et al. (2002) investigate the effect of order crossover. Johnson et al. (1995) discuss the error in fill-rate expressions in periodic inventory systems as a result of double counting of backorders. This only occurs when the lead time is longer than one period and the replenishment order is not sufficient to cover the existing backlogged demand.

The objective here is to investigate how the length of the contract horizon, the demand distribution, and the cost of failing to meet the target (or equivalently the desired probability of achieving the target) affect the fill rate that a supplier achieves over a finite horizon. In §2 I formally describe the inventory system and introduce notation. In §3 I investigate the behavior of the random variable and the implications of this finite horizon behavior on stocking policies. Conclusions and future areas of research are discussed in §4.

2. Model and Notation

To focus solely on the horizon effect, I study a simple, periodic review model with i.i.d. demand, zero lead time, and no ordering cost. It is well known that a base-stock policy will minimize holding and shortage costs when the horizon is infinite (see Zipkin 2000 for a discussion of periodic review inventory models). For a supplier seeking to meet or exceed a fill-rate target over a finite horizon, a stationary stocking policy is not necessarily optimal, however, because such policies are common in practice, I restrict the study here to order-up-to policies.

The model operates as follows: s units are on hand at the beginning of period t and demand is observed and satisfied; if demand exceeds the stock level, excess demand is backlogged and finally a replenishment order is placed to bring the next period's beginning stock level back up to s. Let X_t denote the demand random variables (assumed to be nonnegative), and let $Y_t \equiv \min(s, X_t)$ denote the filled demand. The fill rate random variable for T periods is

$$\alpha_T(s) \equiv \frac{Y_1 + \cdots + Y_T}{X_1 + \cdots + X_T}, \tag{1}$$

which is the infinite-horizon expression for fill rate for i.i.d.; X_t is equal to expected units filled per period divided by expected demand per period,

$$\lim_{T\to\infty} \mathbb{E}\alpha_T(s) = \frac{\mathbb{E}Y}{\mathbb{E}X}. \tag{2}$$

Chen et al. (2003) establish results addressing the *expected* fill rate for a finite horizon. In particular, they show that the expected finite-horizon fill rate is greater than the infinite-horizon fill rate shown in Equation (2) and less than the single-period expected fill rate:

$$\mathbb{E}\alpha_1(s) \geq \mathbb{E}\alpha_T(s) \geq \lim_{T\to\infty} \mathbb{E}\alpha_T(s). \tag{3}$$

The interested reader is referred to Chen (2003) for complete proof details. The intuition behind the finite horizon results in Equation (3) is that in computing the expectation for a finite horizon, the fill-rate realization for a sample path is weighted by the probability of that sample path occurring, but not by the total demand for that sample path. Consider the case where demand in each period is either 1 or 2 with equal probability. With a stock level of 1, the number of units filled per period is always 1, and the expected demand per period is 1.5 for a long run fill rate of 2/3. In any single period, the fill rate is either 1/2 or 1 (equally likely) for an expected single-period fill rate of 3/4. In computing the long-run fill rate in this equally likely case, the 1/2 fill-rate outcome should be given twice the weight of the 1 outcome because demand is twice as large in that case, leading to the long-run fill rate value of 2/3.

In essence, large demand realizations have a smaller effect on the finite horizon expected fill rate than on the long-run fill rate because there is a chance that the supplier will not see large demand realizations within the review horizon. This finite-horizon effect can be quite dramatic as the following extreme example shows. Suppose demand is 1 w.p. $1 - 10^{-k}$ and 10^{2k} w.p. 10^{-k}. With a stock level of 1, the number of units filled in each period will be 1 and the realized fill rate in any period will be either 1 or 10^{-2k}. For large values of k, the expected one-period fill rate is close to 1, while the expected demand per period is close to 10^k, implying that the long-run fill rate is close to 0. In this case, a supplier facing any one-period, service-level contract would meet any target level w.p. $1-10^{-k}$ but have a long-run fill rate of close to 0. In fact, for large enough k, the expected fill rate for any finite-length review horizon will be close to 1 even though the long-run fill rate will be 0.

This phenomenon suggests that a supplier facing a service-level contract might prefer a shorter horizon to a longer one, especially when the per-period demand distribution has a thick right tail. While this is true in some situations, there are countervailing advantages to the supplier of longer review horizons as well. In particular, the potential advantage of a long horizon is that even if the supplier does see large demand realizations, the long horizon may give her enough time to make up for this event and still meet her fill-rate target. Therefore, even though the *expected* fill rate over a finite horizon is greater than the long-run horizon (Equation (3)), an inventory manager may choose to set higher stock levels when facing a finite-horizon performance review. This is explored in greater detail below.

3. Fill-Rate Distribution and Stocking Implications

In this section I examine the effects that per-period demand variability, stock level, and horizon length have on the distribution of α_T. As explicitly calculating values associated with α_T appears to be difficult, I estimate the values reported here using Monte Carlo simulation.[1] Note that the horizon length here corresponds to the number of stocking decisions and could take on a wide range of values. For example, weekly stocking decisions with a monthly performance review would lead to $T=4$, while daily stocking decisions with a quarterly review would give $T=90$.

3.1. Behavior of the Distribution

Figure 1 shows the probability of α_T falling in 1% wide intervals with Erlang (5, 1) demand, a stocking decision based on a long-run fill rate of 95% ($s \approx 7.26$) and review horizon length of $T = 5, 10, 20, 100$. (Descriptive statistics for Figures 1 and 2 are shown in Table 1.) A vertical line denotes the expected fill rate for each case. As noted above, short review horizons have the benefit to the supplier that the expected finite-horizon fill rate is greater than the long-run fill rate. Here we observe, not surprisingly, that with very short horizons, very high and very low realized fill rates are more likely. For example, with $T=5$, there is a 44% ($=F(s)^5$) chance that all demand is satisfied, as well as approximately a 19% chance that the fill rate will be below 90%. For $T = 100$, the probability that the achieved fill rate is 100% is near zero, while the probability of it being less than 90% is approximately 0.16%.

Figure 2 shows fill-rate distributions for $T=20$, with different per-period demand distributions (Erlang 1, 3, 5, 9). All stock levels are based on a long-run fill rate of 95% ($s \approx 3.00, 5.18, 7.26, 11.29$). For larger Erlang "shape" parameter, the per-period demand distribution has lower coefficient of variation and lower positive skewness. Note in comparing Figures 1 and 2 that increasing the value of the shape parameter has a similar effect to increasing the period length. This is not surprising because as noted above, larger shape parameter means lower coefficient of variation and thus lower variability in the achieved fill rate. Similarly, a longer review horizon leads to lower variability in the achieved fill rate.

3.2. Implications on Stocking Decisions

For a manager facing a service-level contract, the elements to be negotiated could include the target level, the length of the review period, and the magnitude of the penalty. Clearly a manager would prefer the lowest target and the smallest penalty. It is not clear, however, whether a manager would prefer a short or long review horizon. As noted above in the discussion of Figure 1, a manager using the 95% long-run stock level would prefer a short horizon if the target level was 100% but a long horizon for a lower target level of 90%.

Of course, managers would choose different stock levels for different contractual commitments.

[1] All simulations use the Mersenne Twister pseudo-random number generator (see Matsumoto and Nishimura 1998) and 10^7 replicates. Error bars would be shown on the graphs, but given the 640×480 resolution of the plots, the half-width of 95% confidence intervals is less than one pixel.

Figure 1 Distributions of Fill Rate with the Same Stocking Level and Per-Period Demand Distribution but Different Horizon Lengths, T

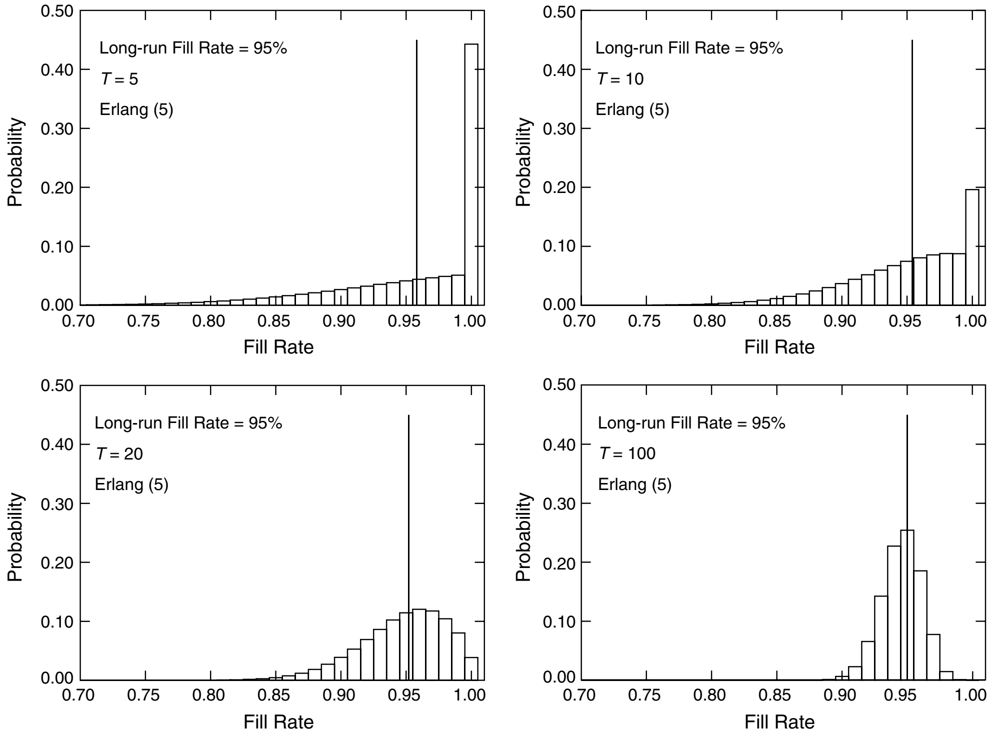


Figure 2 Distributions of Fill Rate with the Same Horizon Lengths (T) and Stocking Levels but Different Per-Period Demand Distributions

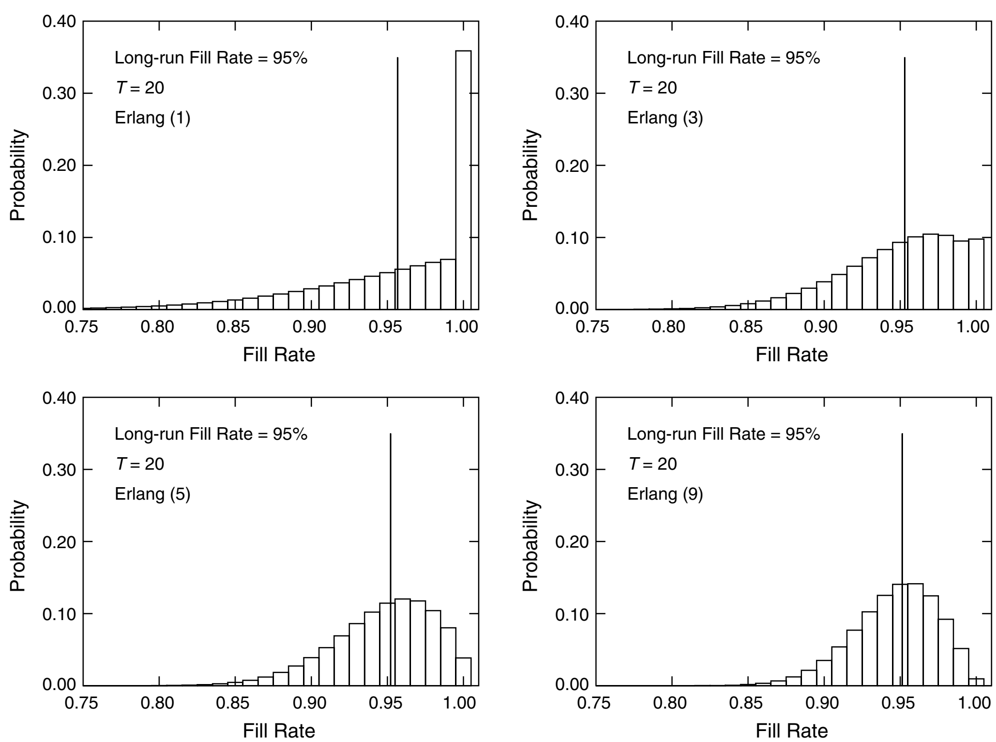


Table 1 Descriptive Statistics for Distributions Shown in Figures 1 and 2

m-Erlang	Long-run fill rate (%)	Horizon length *T*	Mean	Median	Std. deviation	Skewness
5	95	**5**	0.9578	0.9894	0.0585	−1.5955
5	95	**10**	0.9541	0.9647	0.0448	−1.0908
5	95	**20**	0.9521	0.9566	0.0330	−0.7509
5	95	**100**	0.9504	0.9514	0.0153	−0.3262
1	95	20	0.9567	0.9793	0.0550	−1.5416
3	95	20	0.9530	0.9599	0.0388	−0.9339
5	95	20	0.9521	0.9566	0.0330	−0.7509
9	95	20	0.9514	0.9542	0.0276	−0.5934

In choosing the appropriate probability of successfully meeting a target level, a manager would trade off inventory cost versus the penalty cost of failing to meet the target. Figures 3 and 4 show the stock levels necessary to achieve a specific probability of successfully meeting or exceeding a target level for multiple review horizon values. A horizontal line on each graph shows the stock level necessary to achieve the target level in the long run. For all cases shown in these figures, the necessary stock levels increase initially with the review horizon and then eventually decrease; although for the top left graph in Figure 4, the decrease occurs beyond the range of the plot. Comparing plots within each figure, we see that the stock levels decrease faster with the horizon length for lower relative demand variability. This observation is consistent with our discussion above that the short-horizon benefit of potentially avoiding large demand realizations is greater for higher variability demand. Also note that the long-horizon benefit is more pronounced for lower variability demand. For example, all the curves in the lower right in Figure 3 (95% target, 9-Erlang demand) are decreasing after the first five periods.

In comparing analogous plots in Figures 3 and 4, note that for the higher target level (99% in Figure 4), the point at which lengthening the horizon reduces the necessary stock level occurs much later. Recall that the benefit of a long horizon is that a supplier is not doomed by a single, large demand realization. With a very high target fill rate, it simply takes longer to recover.

Finally, note that the required stock levels for many of the cases shown are well above the stock levels necessary to achieve the target level in the long run

Figure 3 Stock Levels Required to Have a Specified Chance of Meeting a 95% Target Fill Rate

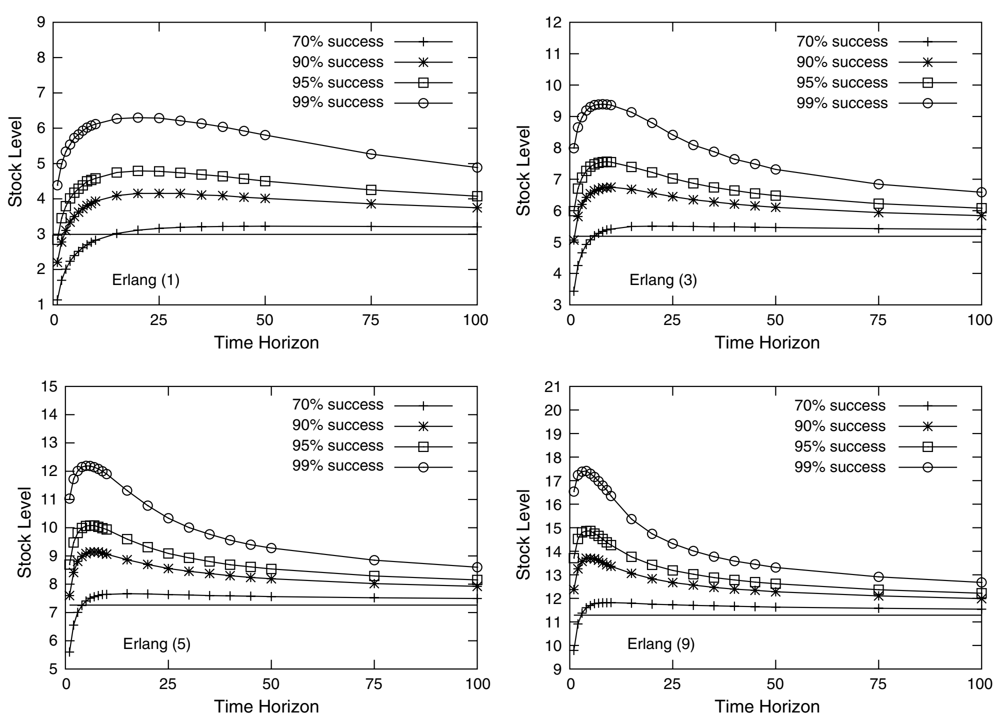
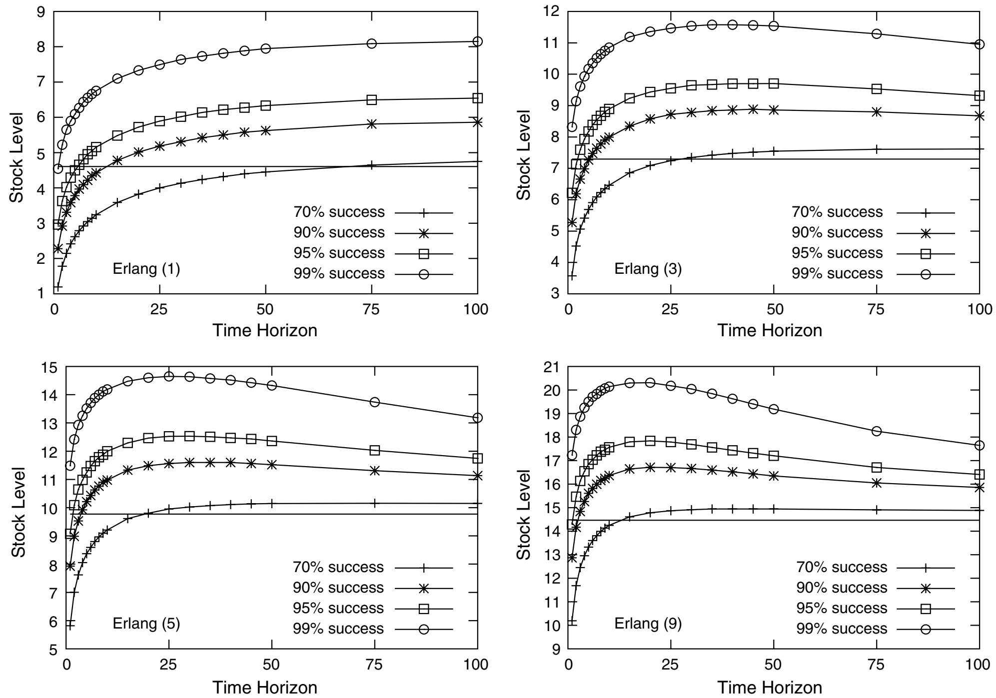


Figure 4 Stock Levels Required to Have a Specified Chance of Meeting a 99% Target Fill Rate



and that the review horizon length can significantly impact the required stock level. For example, in the bottom left graph in Figure 3 (95% target, 5-Erlang demand), a supplier facing a review horizon of $T = 10$ would need to stock over 9.9 units to have a 95% chance of meeting the 95% fill-rate target. This stock level is about 37% higher than the level needed to achieve 95% long-run fill rate and would actually achieve over 99% long-run fill rate. With a review horizon of $T = 25$, a stock level of 9.1 is sufficient to achieve the target 95% of the time (producing 98.4% long-run fill rate), and with $T = 100$, a stock level of 8.1 units achieves the target 95% of the time (producing a 97% long-run fill rate). Also, while in some cases the required "overshoot" is quite large, there are cases where a supplier can actually undershoot, particularly with short horizons, highly variable demand, and a low to moderate chance of success (see 70% success graphs in Figure 4).

4. Conclusions and Future Research

This work was motivated by observing finite-horizon, fill-rate performance contracts in practice. An inventory manager facing such a contract may be able to influence target levels, penalties, and the length of the review horizon. The discussion and examples presented here suggest that these factors can have a significant effect on a supplier's probability of achieving a target fill rate over a fixed review horizon. Short review horizons provide the benefit to the supplier that large demand realizations may not occur during a particular review horizon, increasing the chance that the target fill rate is met. This effect is particularly strong when demand is highly variable and positively skewed and when the target fill rate is very high. Long review horizons increase the chance that large demand realizations are seen but also give the supplier more opportunity to recover from large realizations.

It is important for both suppliers and customers to understand that imposing a finite-horizon, service-level contract may cause a supplier to achieve a long-run fill rate well above the contractually specified target. Recall the example from the previous section where a supplier wanting a 95% chance of achieving a 95% target level would need to set her stock level to achieve a 99% long-run rate with $T = 10$, 98.4% with $T = 25$ and 97% with $T = 100$. In essence, a customer imposing a 95% fill-rate target with a large penalty is really demanding a long-run, fill-rate target substantially larger than the contractually specified level.

I examined a zero lead time, stationary demand, base-stock model with a single customer. The computational work used Erlang demand distributions. One could certainly investigate the effect of different demand distributions that may be nonstationary and/or correlated. The impact of positive lead times could also be investigated. Furthermore, we restricted the inventory policy to be of the order-up-to type. A manager facing service-level contracts with multiple customers could dynamically adjust stocking levels, strategically allocate inventory to customer demands, or both. While it may be too cumbersome for an inventory manager to compute and/or implement optimal policies in practice, it would seem worthwhile to investigate the marginal value of the optimal policy as compared to the best stationary policy.

Acknowledgments
The author is grateful to former Deputy Editor Paul Zipkin as well as the senior editor and two anonymous referees for their insightful comments and feedback on this paper. The author also thanks David Kelton for his suggestions regarding simulation methodology.

References
Chen, J., D. K. Lin, D. J. Thomas. 2003. On the item fill rate for a finite horizon. *Oper. Res. Lett.* **31** 199–123.

Johnson, M. E., H. L. Lee, T. Davis, R. Hall. 1995. Expressions for fill rates in periodic review inventory systems. *Naval Res. Logist.* **42** 57–80.

Law, A. M., W. D. Kelton. 2000. *Simulation Modeling and Analysis*. McGraw-Hill, New York.

Matsumoto, M., T. Nishimura. 1998. Mersenne twister: A 623-dimensionally equidistributed uniform pseudorandom number generator. *ACM Trans. Modeling Comput. Systems* **8** 3–30.

Robinson, L. W., J. R. Bradley, L. J. Thomas. 2002. Consequences of order crossover in inventory replenishment systems. *Manufacturing Service Oper. Management* **3** 175–188.

Tyworth, J. E., L. O'Neill. 1997. Robustness of the normal approximation of lead-time demand in a distribution setting. *Naval Res. Logist.* **44** 164–186.

Zipkin, P. H. 2000. *Foundations of Inventory Management*. McGraw-Hill, New York.