A half century ago, several scholars, some of whom subsequently received Nobel prizes in economics, developed inventory models primarily in response to the needs of American military service branches and a few large corporations. Significant government waste was attributed to mismanagement of weapon systems assets. There were readiness-debilitating shortages in the presence of available supplies of weapon components that were crisscrossing the globe. Big corporations also recognized a potential for profit improvement from getting more bang from a buck of inventory investment. Given that management information systems 50 years ago employed punched-card processing in most organizations, it was impossible for a typical company to adopt the new mathematical inventory models.

Fast forward to year 2002. Today we are frustrated when we repeatedly find a favorite brand out of stock at the local grocery or drug store. We are dismayed at experiencing third-world customer service levels only a few blocks from home. We surmise that the stock shortages are rooted in poor management and not in poor systems.

This essay focuses on why we continue to find empty shelves where we do business despite a half century of impressive research in inventory modeling, augmented by high-priced multiplatform supply-chain management software. We observe companies that have poor customer service despite excessive inventories. Good inventory modeling and advanced inventory control systems are supposed to eliminate such things.

Our story unfolds as follows: The central theme of this reflection is inventory theory in the service of practice—not theory for the sake of theory. First, I will review several formative research findings published between 1950 and 1965. I will make a case for why these research contributions remain relevant today. The references at the end of the article are publications that appear no later than 1965, with only a few exceptions.

Next I will segue to an arguable proposition (with all deference to Ron Howard) that nothing is as impractical as a good theory of inventory. I will explain why what has been truly good inventory research for more than 35 years has not done much to advance the practice of industrial inventory control.

Finally, I will suggest that developments in information technology renew the opportunity to improve practice provided that particular avenues of inventory systems research are pursued.

Silver et al. (1998) published their third edition of an extraordinary text dealing with inventory management. It cites and thematically organizes the findings of more than 1,600 researchers—its topical coverage is encyclopedic (after 1965). In my opinion, it reaches a superlative level of achievement, notwithstanding any of my comments below, and it is readily accessible. Consequently, I am going to refer to it occasionally to back up my propositions; I will use the abbreviation SPP3 whenever I refer to the book.

1. IN SEARCH OF FULFILLMENT

The earliest publications on inventory modeling date back to the 1920s—the lot size (square root EOQ) model is the most notable example, with its commercial context of stocks held in businesses. Before 1950, macroeconomists also wrote about fluctuations in inventory levels in the context of classic business cycles.

In the early 1950s, a few influential research contributions emerged from this primordial condition. Arrow et al. (1951) analyzed probabilistic inventory models. Simon (1952) wrote about servo theory applied to production. Dvoretzky et al. (1952a, 1952b) discussed inventory models and statistical processes. All of these contributions were published in *Econometrica*. Whitin (1953) published a book that was devoted solely to inventory themes, and Bellman, at the RAND Corporation, wrote a monograph on dynamic programming in 1953; see Bellman (1957). The United States Air Force, Navy, and Army funded research efforts aimed at improving the performance of logistics systems. The research impetus driving inventory modeling was in full force by 1954.

In the early 1950s, a few influential research contributions emerged from this primordial condition. Arrow et al. (1951) analyzed probabilistic inventory models. Simon (1952) wrote about servo theory applied to production. Dvoretzky et al. (1952a, 1952b) discussed inventory models and statistical processes. All of these contributions were published in *Econometrica*. Whitin (1953) published a book that was devoted solely to inventory themes, and Bellman, at the RAND Corporation, wrote a monograph on dynamic programming in 1953; see Bellman (1957). The United States Air Force, Navy, and Army funded research efforts aimed at improving the performance of logistics systems. The research impetus driving inventory modeling was in full force by 1954.

These scholarly endeavors encompassed probability modeling (especially renewal and queuing processes), feedback systems, statistical decision theory, microeconomics, and multiperiod optimization. All of these themes are lively today and continue to influence advanced inventory research.
In the early 1950s, it was difficult to obtain historical demand data for individual items (even for weapons systems components stocked at military bases), and computing capacity to automate replenishment formulas was extremely limited. There was little opportunity to empirically test emergent inventory theory.

There were, however, fundamental insights from these early research publications. The new literature revealed the appropriate form (architecture) of replenishment rules, the tractability of discrete vis-à-vis continuous time modeling, the interdependence between the reorder point and reorder quantity, the convenience of particular demand distributions (Poisson, exponential, and normal), and the usefulness of the principle of optimality. Since there was little available empirical demand data at the level of a stockkeeping unit (SKU), assuming a Poisson or exponential distribution versus a normal distribution was a matter of mathematical elegance versus flexibility in being able to specify both a mean and standard deviation. Square-root-type formulas using continuous time models were computationally more practical than renewal recursions. And mathematically derived replenishment policies that explicitly accounted for imprecision associated with observational data seemed too elegant, if not esoteric, to implement.

Early 1950s research also shed light on technical details regarding felicitous model formulations. It was helpful to assume that the economic impact of customer service level is included in the model’s objective function to be optimized (rather than to express it as a side condition). Unmet demand is fully backlogged, lead time is knowable and deterministic, demand is iid for an individual SKU (which does not deteriorate or become obsolete), the time horizon is a single period or unbounded, and if the latter, all parameter values are stationary. Even what seems to be slight departures from these assumptions were known to create serious analytic challenges.

The aforementioned stochastic inventory modeling research was complemented by deterministic multiperiod multiproduct linear-programming formulations. In the early 1950s, however, performing linear-programming optimization (by the simplex method) meant using punch-card computers. By necessity these models were toy-sized. Only later did it become feasible to think about deterministic multiperiod linear-programming models as a way to consider multiproduct time-phased production and inventory decisions.

An important legacy of the early 1950s is an insight about the architecture of inventory control solutions. Optimizing dynamic stochastic models implies finding a strategy or policy—in other words, a rule in which all future decisions are contingent on future states of the system. These states are determined in part by the future outcomes of random events. Consequently, future decisions are described probabilistically. In contrast, optimizing deterministic linear-programming models implies using an algorithm that yields numeric values for all future decisions—therefore all of these values are knowable at the outset of the planning horizon. These subtleties are taken for granted today.

Inventory research is patently indebted, even now, to the path-breaking modeling of the early 1950s.

2. POINTS OF LIGHT

For a decade beginning in the mid-1950s, the core ideas above were pursued vigorously. With the notable exception of Brown (1959), publications assumed that either future demand values are given or that their underlying distribution is completely known. Most researchers continue to make these assumptions. I will summarize a few of the pivotal themes from this ten-year span.

Wagner and Whitin (1957, 1958), and Manne (1958) extended the classic deterministic lot-size model with stationary demand to accommodate known demand that fluctuates from period to period. The dynamic lot-size approach was eventually recognized to be a deterministic version of a renewal model, and equivalent to finding a shortest route in an acyclic network. Nevertheless, even sophisticated texts like SPP3 explain it in a complicated way, despite the fact that the dynamic lot-size model is a much simpler acyclic network than a typical critical path, which is standard material in OM textbooks. The network characterization relies on the concave property of the objective function. Bowman (1956) and Johnson (1957) explored alternative convex objective function formulations, and established the consequent optimality of a myopic decision process. The idea is closely related to the notion of a greedy solution process (optimize incrementally and never revise a prior decision).

The concept of a finite planning horizon determined by the data emerged from the lot-size modeling above. Conditions were discovered that ensured an optimal solution for a finite horizon remained optimal for an extended horizon; further, these conditions did not require full knowledge about the longer horizon.

Complementary to renewal theory research, Morse (1958) examined steady-state stochastic replenishment systems using queuing theory. He viewed inventory items as analogous to servers in a queuing system, server busy time as tantamount to replenishment lead time, and a waiting line as comparable to customer demand backlog.

The mathematical form of optimal multiperiod decision rules in the presence of stochastic demand was explored in different ways. Holt et al. (1956, 1960) demonstrated that under certain assumptions, an optimal rule is linear in the parameters of the demand distribution. These assumptions, however, did not turn out to be sufficiently appealing to motivate much further research.

It was known by the mid-1950s that the form of an optimal policy is sensitive to whether the objective function contains a setup cost and a smooth cost function associated with less than perfect service, the delivery lead time is greater than one period, and unfilled demand is fully backlogged. The noteworthy research achievements mentioned next address these technical challenges.
Using reasonable assumptions, Scarf (1960) established the optimality of \((s, S)\) policies: When inventory on hand plus inventory due in less backlog falls below \(s\), order enough to bring it up to \(S\) — else, do not order. (It was known that under alternative plausible model assumptions, this simple policy was not optimal.) Roberts (1962) provided a way to compute approximately optimal policies. Veinott and Wagner (1965), and subsequently other researchers, investigated effective computational methods for obtaining optimal policies, and later, approximately optimal policies.

Veinott (1965) showed for an important class of situations that a myopic replenishment policy is optimal, and thereby established that an unbounded horizon model can be decoupled so as to yield near-term optimal decisions using only near-term parameter values. Further, the work established a sound basis for what has become an important decision rule in practice: Replenish what you sell. This is an elementary example of a so-called pull system.

Clark and Scarf (1960) formulated a seminal and tractable model for inventory replenishment in an environment comprised of several echelons that hold inventory and where final demand is uncertain.

Operations research textbooks at that time did not make sharp distinctions among different inventory management settings. Inventory scholars realized by then, however, that inventory theory approximates reality well when a stocked item is ordered from an outside vendor, but not so well when an item is manufactured by the enterprise as a direct result of a replenishment decision. Today this underlying distinction is evident and reflected in finite capacity scheduling models. Expository habits are hard to break, nevertheless, and even SPP3 discusses EOQ in terms that allow for both interpretations of purchasing from an outside vendor and manufacturing the SKU from within the enterprise.

By the early 1960s, it was clear that linear programming is a conceptually well-suited alternative for addressing a combination of multi-item, multilocation, multiperiod issues in a capacity-constrained environment, even given its limitations. In contrast, it was hard to imagine that anything practical would result from an extension of stochastic inventory theory models in a pull context to multi-item situations (that is, to something beyond applying single-item analysis to each SKU individually). The reason is that it would be a heroic task to use historic data to establish multivariate demand distributions.

The fundamental distinction between multi-item and single-item models corresponds closely to the split in practice between push and pull inventory control systems. In a multi-item manufacturing environment, pull systems, which aggregate orders for all items that are requested, may lead to infeasible production schedules as well as periods with excess capacity; thus push systems tend to be the rule in practice. In a stock-replenishment environment, push systems, which keep capacity fully utilized, may lead to overstocking as well as excessive obsolescence; thus pull systems tend to be the rule in practice, although transportation constraints sometimes countervail.

Forrester (1958, 1961) published an article and later a book on what he called industrial dynamics, and thereby created a computational engine that exemplifies economists’ traditional business-cycle logic. The approach views a SKU’s inventory level as a time series mathematically created by the difference between cumulative production and cumulative demand, layered on a base of safety stock. The industrial dynamics model facilitates visualizing the imposition of boundary conditions on the time series (and their slopes), along with the imposition of servomechanisms (feedback). Despite industrial dynamics’ broad sweep, it does not adequately address probabilistic uncertainty, which ultimately has limited its contribution to inventory control systems (in particular, how to set safety stocks).

Wagner (1962) published a research monograph under the Operations Research Society’s sponsorship that introduced statistical issues of importance to senior management. The research theme focuses on how corporate management can use aggregate economic measures to ascertain whether lower echelons of a supply chain are adhering to automated inventory replenishment logic. Recognizing statistical uncertainty is essential in assessing the aggregate measures.

By the mid-1960s, inventory modeling was technically sophisticated. The easy (and some not-so-easy) wins were already won. Most of the restrictive assumptions used in the prior 10 years had been relaxed at least in exploratory research. In the decades to follow, the technical horizons continued to expand, although the managerial scope of the theory remained much the same. Inventory theory analyses rarely merged with strategic management deliberations. Analytic inventory models usually take as given what strategic thinking views as choices. For example, product line breadth, a strategic choice but implicitly given in inventory modeling, influences demand for stocked items.

### 3. SHOW ME THE DATA

By the early 1950s, the theory and practice of inventory control faced a fundamental issue: The specification of demand and lead-time processes cannot be done with much precision. It was presumed, however, that once it was possible to obtain timely historic data on a continuing basis, sophisticated replenishment formulas could be easily applied by using appropriate statistical methods.

Brown’s *Statistical Forecasting for Inventory Control* (1959) was groundbreaking. He suggested specific statistical methods to use, particularly the approach that he named as exponential smoothing. (In 1956, Brown presented his ideas at an ORSA conference, and in 1957, Holt wrote an Office of Naval Research report discussing exponential weighted moving averages.) In his book, Brown illustrates the calculations with hand-drawn worksheets. The book appears to encompass both manual systems and automated calculations; clearly, this was a transitional moment.
with respect to computing power. Brown points out that exponential smoothing does not require keeping long files of prior demands, viewed as an advantage at that time but irrelevant, if not legally impossible, today.

Brown expected his readers to know the fundamentals of inventory control. It is not easy, however, to spot any fully developed mathematical replenishment models in Brown’s book. But Brown does set forth a stock-replenishment point of view by posing two central questions in automated inventory control

(a) Is it now time to replenish inventory?
(b) What should be the order quantity?

His approach to these questions is rooted in the requirements of practice and not in the underpinnings of inventory theory. This is evident from the way the questions are posed—the operative word in (a) is now. We review the implied architecture below.

To make the replenishment-order decision (a) above, assume that replenishment review occurs at the start of each period. Practical considerations rise when we attempt to ascertain a SKU’s current stock-status level, defined as inventory on hand plus inventory due in, less backlog. Frequently, the stock-status level is hard to calculate, given the organization of corporate MIS systems—typically the components of stock status are in different transaction systems. We determine whether the stock-status level gives enough service protection for demand over lead time plus a review period, taking into account uncertainty in demand and perhaps lead time. We order now if the stock-status level does not provide enough service protection; else, we wait for another period and test again.

One of two situations arises. If bona fide probability distributions for the uncertain quantities are available, the above determination about service can be made using a mathematical inventory model. We directly apply a formula to calculate $s$ (reorder point). But if there is only historic data about these quantities, then we use the data to forecast demand over lead time plus a review period, and prospectively take account of accompanying forecast error so as to hedge against demand and lead-time uncertainties. The hedge often is called safety stock, and how to calculate it is the challenge.

A similar dichotomy occurs in answering the replenishment-quantity decision (b) above. If the relevant probability distributions are known, we calculate $S$ (the order-up-to point), and order enough to bring the stock-status level to $S$. But if there is only historic data, the alternative is to forecast what will be sold over a span of time that is expected to elapse until the next replenishment, possibly adjusting the amount upward if the stock-status level is exceptionally low.

A statistical approach is a far cry from operations research logic, such as that underlying a classic $(s, S)$ policy. You can spot the difference immediately by observing that a statistical approach produces two numbers based on history as of now, whereas an OR model produces a reorder policy that is applicable (possibly) over an unbounded horizon. An OR model can be implemented in practice by using repeatedly updated statistical estimates of the parameter values in an assumed demand probability distribution.

In the discussion that follows, I designate the approach of using point forecasts and observed variation in forecast error by PFErr. I designate the alternative approach of using an OR inventory model populated with statistically estimated parameters of an assumed demand distribution by OReSt. (You will find it helpful to jot down the definitions of these two ad hoc abbreviations.) What we want to explore further is whether one of the two approaches is significantly more effective than the other in practice.

It is unfortunate that the term forecast has been identified with the data analysis suggested by Brown and others. To a manager, a forecast refers to the demand value that actually will be observed; since it is hard to perfectly anticipate future sales, a point forecast is not to be believed (unless one has a crystal ball). The output of exponential smoothing and similar calculations is an estimate of mean future demand, which is a concept—the mean itself is never actually observed. Likewise, the term forecast error is interpreted ex post by managers as a mistake, possibly a misjudgment, whereas it is only one observation from a distribution of forecast errors. The distribution purportedly reveals inherent uncertainty about future demand and can be used to determine a value for safety stock. Academic readers may find this niggling over terminology only mildly amusing, but the misunderstanding of this terminology in common use has been the downfall of many practitioners. Managers do not grasp what they are going to get.

Brown’s point of view is reflected in today’s commercial supply-chain software packages: Historical data are utilized to make point forecasts of future demand. Further, uncertainty in future demand is formulated by an estimate of the standard deviation of forecast error; lead-time uncertainty is finessed by some other approximation (I will explain further below). Practitioners bought into the statistical point of view right away, whereas most inventory theorists gave it short shrift.

The chapter organization in the book SPP3 lends support for the preceding discussion. Most of SPP3 is devoted to mathematical models, that is, the results of some 1,600 theorists; the cited work provides decision-support tools in supply chains. With few exceptions, these models assume that the demand and lead-time uncertainties are described completely by known (or postulated) probability distributions with precisely specified parameter values. SPP3 makes it quite clear in a single long and comprehensive chapter devoted to forecasting that a supply chain is often comprised of nonstationary uncertainties. Nowhere is there an explicit consideration of the impact of statistical noise (from having only finite data from a nonstationary environment) on the performance of the mathematical models. Yet dealing with this impact is one of the major challenges facing a practitioner.
4. IT’S THE DATA’S FAULT

With the passage of 50 years, one thing is true by now. There is no reason to complain anymore about scarcity of data. All businesses in the United States have loads of sales data. Usually this data is accessible for at least two years back, and often further back than that (although the earlier data may not be online).

The agony that is felt today reflects that the demand data are dirty. This is not quite the same as what statisticians call missing observations and errors of measurement, although dirty data may include both.

Here are some typical examples of data challenges.
- Demand is not level across a year, so a twelve-week running average of demand fluctuates considerably.
- A SKU is new and being phased in, or is old, obsolete, uncompetitive and being phased out; in either case, the span of historical data is limited.
- A SKU was new last year and introduced without enough supply to meet demand, so last year’s sales understated potential demand.
- A supplier failed to ship orders and consequently the SKU was out of stock, so historic sales understated demand for a span and overstate demand in the weeks immediately after.
- A SKU received special promotions (maybe under a different SKU number, and with a different set of promotion dates, depending on the physical location of demand), so weekly sales data show spikes and subsequent valleys, and these differ by location.
- Several times last year, demand was unusually large due to the unexpected orders of a few large customers, so the mean and standard deviation of historic demand may not be appropriate for inventory control.
- The SKU’s product specifications changed (color, flavor, package size, quality, etc.), so there is discontinuity in the pattern of historic demand.
- Customers’ returns are netted out of a current week’s sales, so raw sales may overstate demand in one time period and understate it later.
- Some holidays fall on different days and weeks from one year to the next, so sales data have to be adjusted from one year to the next to obtain comparability.
- Weather impacts demand, so historic sales reflect special conditions that may not recur.
- A competitor introduced a new product that altered market share, so historic demand declined—the size of the reduction varies, reflecting the number of weeks since the new product introduction.
- A new SKU was introduced and thereby cannibalized the sales of several other SKUs.

This list is long enough to convey why the phrase garbage-in-garbage-out is often used in discussions about demand forecasting and the distribution of forecast errors, and why automated replenishment systems commonly under-perform relative to manager’s expectations.

The preceding examples give rise to distinct classes of data problems. Some of the issues cited are solvable by suturing data streams; this involves identifying holidays, promotional weeks, and other special events. But other data issues are fundamental and reflect business practices—an important example is when a SKU’s life span is less than two years.

Some companies have invested in computer systems that assist in cleaning up dirty data. (These are almost never off-the-shelf software packages; the system requires considerable hands-on guidance.)

Consider a company that has an effective MIS system so that the cleaned sales data reasonably state what the relevant historic demand is for each SKU. This setting makes it easier to compare alternative architectures for automated inventory control. Note in passing that systems for cleaning historic data rarely clean previous forecasts. Hence, automated replenishment logic based on the distribution of historical forecast error is problematic when data cleaning is implemented. This is another reason to be circumspect about adopting an automated replenishment architecture that relies on forecast errors.

It is unlikely that in practice there is much difference between the effectiveness of PFErr and OREst in deciding what should be the order quantity? Here is the reason. Suppose that the real situation is so simple that the stationary EOQ formula applies. Then early research established that EOQ is often a good approximation to the optimal value of $S - s$ in an $(s, S)$ model. Further, the actual value of the objective function is insensitive to an order quantity that differs considerably from EOQ. In most real situations, however, there are other factors that influence the order quantity, such as pack size, discount pricing, transportation minimums and maximums, etc.

The classic lot-size model is formulated to characterize the order decision as a quantity of goods. An equivalent alternative is to characterize the decision as a time interval over which the imminent order should last, such as weeks of coverage. This perspective yields an EOI (economic order interval). Point forecasts of future demand over EOI can be effective and integrated equally as well into a PFErr or OREst system. Note that when the EOI is a single period, the order-quantity decision comes down to how close the replenishment quantity should be to what was sold last period.

An analytic challenge with respect to order quantity is recognizing in advance when to make the final order for a SKU, and then choosing how much. It always is clearer in retrospect which order is responsible for unsalable leftover stock.

Our comparison of PFErr and OREst systems rests primarily on how to determine is it now time to replenish inventory? Given the current stock-status level, the question is whether postponing a reorder for at least another period gives acceptable customer service over the interval from now until an order placed next period would be delivered. This is a prospective assessment and essentially uses probabilistic thinking. PFErr and OREst differ essentially
in detail, if not concept, as to how to calculate the expected service level associated with not ordering now.

PFErr is implemented in today’s supply-chain software systems as follows.

- First, stock-status level (or some analogous measure of inventory) is converted into an ordinate of a standardized normal distribution by subtracting off the point forecast for demand over lead time plus a review period and dividing by the variability of historic forecast error (usually, the standard deviation of past forecast errors, properly scaled using a square root calculation to provide an interval of lead time plus a review period). A serious complication arises when lead time is uncertain. We discuss this issue later.

- Second, a standardized ordinate of the demand distribution need not be close to zero, mean square error is rarely used in practice. Although mean forecast error need not be close to zero, mean square error is rarely used in practice.

- Third, this standardized expected loss is scaled up to original demand units to provide expected lost demand over lead time plus a review period.

- Finally, expected service is measured as the ratio of expected lost demand to the comparable point forecast of demand.

An alternative articulation of this process works backward from a target service level to an implied critical ordinate of the standardized unit normal loss function, and finally to a reorder level. Looked at this way, all SKUs that have the same target service level have the same critical ordinate. That common ordinate value is a consequence of assuming that forecast errors for a SKU are independently and normally distributed, so that a single inverse function applies.

Missing from this algorithm is a correction that may be required due to forecast bias. Although mean forecast error need not be close to zero, mean square error is rarely used in practice.

If lead time is uncertain, the above calculations for variability become more complicated (and often more difficult to implement because historic data for estimating lead-time variation is limited). If expected lead time is two weeks, but it turns out that an order does not arrive until four weeks, some customer demand in the third and fourth weeks may not be satisfied from inventory on hand.

Whereas the demand distribution for a given lead time may be viewed as concentrated around a single modal value, adding uncertainty in lead time can deform the one-period demand distribution into a composite distribution with several modal values corresponding to a discrete number of weeks of lead time. If targeted service level is high, then safety stock has to cover the rightmost possible values for demand. The composite distribution need not be well approximated by a normal distribution.

You find in practice two different approaches to accommodating uncertain lead time. One assumes that lead time is fixed at a value safely above its mean (such as the 90th percentile). This approach impacts the forecast demand over lead time, but does not affect the value expressing demand forecast uncertainty. The other uses a standard formula comprised of statistical estimates of mean demand, mean lead time, variance of forecast error, and variance of lead time. Typically the statistical estimate for the variance of lead time is seriously in error.

Commercially available supply-chain software that purports to do these calculations sometimes documents the algorithms. But often the software architecture makes it practically impossible for users to verify the underlying logic by testing numeric examples. In simple terms, these systems operate in real time and do not easily accommodate what if explorations. Consequently, when unintuitive forecasts and resupply recommendations occur, the user must take the results on faith.

Suppose that the PFErr system is technically sound. The final step above produces a fraction—for illustrative purposes, assume that the number is .93, which represents 93% service over the relevant time slice. If .93 is below management’s targeted service level for this time span, a replenishment occurs. Otherwise it does not. In any case, the calculations are repeated with updated forecasts and stock status at the start of the next period.

Practitioners naively assume that .93 (in our illustration) is an accurate evaluation of expected service. Rarely, if ever, does such a value give a sufficiently accurate evaluation, and typically it overstates service performance noticeably. No wonder many companies are dissatisfied with the service performance of automatic replenishment software. They expect to get high service at an acceptable level of inventory investment, and actual service falls short of expectation.

To conserve space, I omit a review of all the underlying assumptions in PFErr that in practice are seriously violated and contribute to the illusion that a replenishment system will actually obtain a service level close to what is targeted. The main point in this discussion is that PFErr does not deliver what it promises.

OREst has comparable problems. When OREst employs historic data to estimate the mean and standard deviation of a postulated demand distribution, the calculated service measure may well be an overestimate due to the postulated distribution being inappropriate and the parameter estimates from data being inaccurate. The underlying mathematical reason is that the statistical error in an OR inventory model impacts service asymmetrically and nonlinearly; in other words, the improvement in service from an overestimate of a SKU’s mean or variability of demand is not the same magnitude as the degradation in service from an underestimate by the same amount.

In realistic settings the amount of historic demand information is not sufficient to provide much confidence that the shape of the demand distribution (which is unknown) is well represented with respect to that part of the assumed distribution that is most critical, the right tail.

The above discussion brings us to a realization about our objective of answering is PFErr or OREst significantly more effective in practice? We must posit assumptions about the demand process that gives rise to the historic data.
We seem to have a chicken-or-the-egg conundrum. This logic puzzle is a familiar one to statisticians, who examine analogous issues in data-driven analyses. When is it prudent to assume an underlying normal distribution when making a decision based on limited data?

Determining the appropriate value for safety stock is a serious challenge for practitioners because in today’s business environment, management wants its supply chain to perform at an extremely high level of customer service. At such high service levels, this determination is the hardest to make—the accuracy of data-driven estimates is poor.

5. SIMULATE TO CALIBRATE

There are enlightened practitioners who do have a way of dealing with this issue. I designate this process as model calibration.

These practitioners apply what is called retrospective simulation to a collection (possibly a stratified random sample) of SKUs. They utilize historic data rather than estimated distributions in a simulation of SKU replenishment, where the target service level is made identical for all SKUs in the collection. This yields a simulated measure of service, and is compared to the target service level that is used in the PFE rr or implied by the ORE st formula. As an illustration, a data-driven simulation that targets .98 service overall may result in obtaining .93 service overall from the retrospective simulations. In a subsequent step, practitioners adjust the targeted service amount to get the desired actual service. It is now feasible to perform this calibration process using spreadsheet software on a fast PC.

The consequence of avoiding an assumed probability distribution for demand, such as a normal distribution, becomes clear once a retrospective simulation is performed. A mathematical distribution like the normal has a positive probability of exceeding any given value for the ordinate. Hence, no matter what value is used for safety stock, there is a chance that some demand goes unfulfilled.

In contrast, in a retrospective simulation using finite historic data, there is a value of safety stock for each particular SKU that gives 100% service. As a consequence, as the target service level is raised in the calibration process, more and more SKUs reach 100% service. But other SKUs may not be close to the desired service level, and therefore the overall service level may not be sufficiently high. Raising the target level further provides no improvement for those SKUs already at 100% service, but adds to their inventory. The irony is that all unfilled demand in a retrospective simulation was actually filled historically, since these data represent actual sales, not potential demand.

Consider what happens when calibration becomes an integral part of the systems design process. The calibration mechanics enable a practitioner to ascertain trade-offs between simulated service overall, inventory investment overall, as well as other ancillary measures. Hence, using calibration, it is possible to make direct comparisons of alternative replenishment approaches, such as PFE rr and ORE st, by looking at an aggregation of results from simulating a multitude of SKUs.

The calibration process is a cross-sectional variation on what statisticians call bootstrap analysis. The same replenishment architecture (for example, PFE rr) and target service level are used for a sample of SKUs, where each SKU has its own historic demand.

Calibration is best practice in system design, but it is not standard practice by any means, and not easily accomplished with today’s supply-chain software. For that reason, the typical PFE rr implementation of an automatic inventory replenishment system is fraught with worse-than-expected outcomes, and consequent overrides of the underlying replenishment logic.

6. RESEARCH OPPORTUNITIES

Unlike areas of OR where current research is technically (mathematically) far ahead of practice, current inventory research investigates issues of central importance to practice. Therefore, to the extent that such research reveals important insights about supply-chain architecture, the research expands knowledge in a way that deepens understanding of how these processes behave. Further, the enhanced understanding builds on the established base of prior knowledge.

Recognizing this, it also must be said that incremental mathematical research is not likely to enhance practice. The reason is that mathematical inventory research is blind to all the data issues discussed above and far removed from the entrenched software that now drives supply-chain systems.

The references at the end of this article contain a few examples of research efforts that have dealt with statistical issues. For the most part, these studies assume that demand occurs in a stationary environment, and that lead time is a known value. I am aware of other relevant articles with less restrictive assumptions, but I have not included them in the references.

It seems promising to explore a new research avenue that takes as its starting point a finite history of demand data for a family of SKUs. Unlike the publications of 40 to 50 years ago, the new research should assume plentiful, recognizably dirty, demand data. Lead-time uncertainty needs to be examined from a data perspective as well. The objective should be stock and replenishment logic that is driven by such data, and operating in the context of a supply chain.

The overall research objective is to find sufficiently robust approaches that produce reliable trade-off assessments between service and other aspects of inventory management. The theory should rest on the analysis of a family of SKUs, although the replenishment strategy itself may be single-item focused. The family of SKUs should be defined in a way that its members are recognizable by reference, at least, to historic demand over a finite history.

The theory should encompass data-oriented tests as to when the actual replenishment rules may be reliably applied. The theory should provide measures of accuracy.
with respect to estimates of future systemwide behavior comparable to measures of estimation error commonplace in standard statistics.

The theory should recognize that most demand environments change over time, usually gradually but sometimes abruptly. It should assume that on occasion human intervention is called for, and the theory should assist in signaling when such occasions are warranted. Finally, it should take notice of rules of thumb that may, at first glance, seem too simple (such as replenishment rules based on weeks of supply, or kth largest prior week of demand), but that may function well in a nonstationary environment that contains both demand and lead-time uncertainties.

PERSONAL RECOLLECTIONS

My career interest in inventory and statistical processes was fostered in the beginning by eight scholars, some being mentors, some colleagues, a few both, and all good friends. The pivotal circumstances occurred in four places, Stanford University, RAND Corporation, Massachusetts Institute of Technology, and McKinsey & Co.

Kenneth Arrow and Gerald Lieberman at Stanford wisely counseled me as an undergraduate during the early 1950s about exciting career opportunities in the emerging field of operations research. They opened my eyes about studying economics and mathematical statistics. Until then I never realized that applied mathematics could provide a livelihood. Jerry guided me in learning about important business applications of statistical analyses. Ken assisted me in spending summers at RAND, starting in 1953, and thereby first sparked my interest in inventory modeling.

In 1957, I became a junior colleague of Jerry’s in the Department of Industrial Engineering at Stanford. I had an office in Serra House, in proximity to Ken and his preeminent co-authors. They were actively collaborating on inventory theory, and writing papers that comprised several Stanford University Press research monographs.

In the early 1960s, Arthur (Pete) Veinott, Jr. also joined the faculty in Stanford’s Department of Industrial Engineering. Pete and I shared a deep interest in inventory and production algorithms, resulting in a dialog that lasted for years. Along the way, we developed a practical algorithm to compute (s, S) policies for discrete probability distributions. Pete’s broad perspective on the issues central to inventory and production modeling was inspiring.

In 1955–1957, while doing doctoral studies in mathematical economics at MIT, I was fortunate to be assigned as a research assistant to Thomson Whitin, then on the faculty of the Sloan School of Management. Tom was preparing the second edition of his monograph on inventory processes, and posed the core problem that led to our collaboration on the dynamic economic lot-size model.

Robert Solow, in guiding my doctoral studies, counseled me as to the importance of broadly examining alternative analytic perspectives in researching a scholarly area. I selected for my MIT doctoral thesis topic an applied statistics problem growing out of the research program of RAND’s Logistics Department. The central theme of the thesis was the design of top-level managerial systems for controlling lower echelons in a supply chain. The research applied statistical principles in fashioning a control system comprised of (s, S) inventory policies. Both Ken and Bob gave generously of their time in challenging my thinking about these issues. In 1962 the thesis was published in the Operations Research Society’s research monograph series.

I never had the opportunity to collaborate with William Cooper at Carnegie. But Bill’s dedication to doctoral students was already legendary before the 1960s, and Bill provided a role model for me early on: I am forever grateful.

Robert Fetter, who was on the Sloan School faculty when I was studying at MIT, was interested in practical applications of operations research, and we had many lengthy and wide-ranging conversations about advancements in operations research. It was Bob who introduced me to McKinsey & Co. while I was on the Stanford faculty. Then in 1967, Bob became a close colleague again when I joined him on the faculty at Yale.

My working relationship with McKinsey & Company began in 1960 in San Francisco. Shortly thereafter, David Hertz became a partner at McKinsey in New York. I first met David when he served as editor of the Operations Research Society research monograph series. David kindly tutored me on the skills that are required to take university research and make it relevant to business.

Over four decades, I have had innumerable opportunities to participate with McKinsey client service teams in designing and implementing what is now called supply-chain logistics systems.

In the early 1960s it already was evident that software packages developed by leading computer companies such as IBM were going to be critically important in getting businesses to adopt operations research inventory models. Whereas commercial applications of mathematical programming algorithms seemed, from the 1960s onward, to develop in concert with advances in computing technology, inventory modeling did not fare so well. This article takes a retrospective look as to why.

ACKNOWLEDGMENTS

I am grateful for suggestions made by my colleagues Jack Evans, Geraldo Ferrer, Kyle Catanni, Ann Maruchek, Jayashanker Swaminathan, and Clay Whybark, and the editors Lawrence Wein and Frederic Murphy.

REFERENCES


Kaufman, R. L. 1977. (s, S) inventory policies in a nonstationary demand environment. Working paper, University of North Carolina at Chapel Hill, Chapel Hill, NC.

—. 1976. Multi-item inventory system policies using statistical estimates: Sporadic demand. Working paper, Yale University, New Haven, CT.

Klincewicz, J. G. 1976a. Biased variance estimators for statistical inventory policies. Yale University, New Haven, CT.

—. 1976b. Inventory control using statistical estimates: The power approximation and sporadic demands. Yale University, New Haven, CT.

—. 1976c. The power approximation: Control of multi-item inventory systems with constant standard-deviation-to-mean ratio for demand. Yale University, New Haven, CT.


