Creating an Inventory Hedge for Markov-Modulated Poisson Demand: An Application and Model

Hari S. Abhyankar
PriceWaterhouseCoopers
144 Middlesex Turnpike
Burlington, MA 01803

Stephen C. Graves
Leaders for Manufacturing Program and A. P. Sloan School of Management
Massachusetts Institute of Technology, Cambridge MA 02139-4307

January 2000

Many firms face environments with long production lead times, great product variety, and uncertain, non-stationary demand. A challenge is how to plan production and inventories to provide competitive customer service at least cost. In this paper, we first describe an application at Teradyne in which we implemented an inventory hedge to protect against cyclic demand variability. Based on this experience, we develop a model to understand better the efficacy of this hedging policy. We model an inventory system for a single aggregate product with a Markov-modulated Poisson demand process. We provide approximate performance measures for this system and develop a relevant optimization problem for determining the size and location of an intermediate-decoupling inventory. We use this optimization to show the value of an intermediate-decoupling inventory as a hedge for cyclic demand environments.

Mailing address: MIT, E40-439; 77 Massachusetts Avenue; Cambridge MA 02139-4307. Ph 617 253 6602; Fax 617 253 1462; Email sgraves@mit.edu.

Introduction

Many manufacturers struggle with how best to buffer against uncertainty in demand volume and mix for customized, make-to-order products. Quite often the manufacturing lead-time is much longer than the customer service time. As a consequence, manufacturers must employ tactics for planning materials and production based on forecasts of demand.

In this paper, we examine an instance of this problem in which demand comes from a non-stationary process. In particular, we assume that aggregate demand follows a cyclic pattern in which low-demand intervals alternate with high-demand intervals. We model this demand as a two-state Markov-modulated Poisson process.

We first describe an application at Teradyne that motivates this research. Teradyne is subject to highly variable demand, driven by capital investment cycles in the semi-conductor industry. We helped them to develop and implement a hedging policy, as a planning tactic to address this cyclic variability. In particular, the policy creates a hedge in their master schedule as a means to build a safety stock in their material pipeline. In effect, this hedge acts like an intermediate-decoupling inventory in a multi-stage supply chain.

Based on this application, we posed the research question of how to design and parameterize such a hedging policy. This type of hedging policy is not new. Miller (1979) describes this planning tactic; Wijngaard and Wortmann (1985) provide a more detailed treatment of hedging the master schedule in the context of stationary demand. Guerrero et al. (1986) provide a simulation study of this policy; see also Vollman, Berry and Whybark (1992) and Baker (1993). However, we are not aware of prior research that attempts to model and analyze a hedging policy as a tactic for cyclic variability.

We develop a model to characterize the inventory requirements and service level for a hedging policy for a two-stage system subject to a two-state Markov-modulated Poisson demand process. As such, our work is related to that of Song and Zipkin (1992, 1996) and Chen and Song (1997).

Song and Zipkin (1992) consider a multi-echelon system with a Markov-modulated Poisson demand process. They assume that each stage controls its inventory with a base-stock policy that is independent of the state of the demand process. They develop an exact procedure to characterize the steady-state performance of the inventory system. In their second paper, Song and Zipkin (1996) assume the same demand process for a two-echelon system consisting of a warehouse serving multiple retail sites. The retail sites again operate with an independent base-stock policy, while the warehouse employs a state-dependent base-stock policy. They again develop a procedure to characterize the steady-state performance of the inventory system.

Our work differs from that of Song and Zipkin in that we consider a simpler two-stage serial system, for which the downstream stage has a state-dependent base-stock policy while the upstream stage operates with an independent base-stock policy. Also, the intent of our model is to provide some insight into the behavior and benefit from the inventory hedging policy; as such, we have structured the model to permit the optimization of where to place the inventory hedge.

Chen and Song (1997) consider a multi-stage serial system for which the demand distribution in each period depends upon the state of a Markov chain. They assume linear holding and backorder costs, and show that a state-dependent echelon base-stock policy is optimal for each stage. Our model differs from that of Chen and Song in that we assume a continuous review inventory policy, and service level targets rather than backorder costs.

Nevertheless, their result provides evidence that our inventory policy is reasonable and possibly near optimal, as it maps into a state-dependent echelon base-stock policy.

The remainder of the paper consists of five sections. In the second section we describe the application at Teradyne that introduces and motivates the research problem. In the third section, we develop the model to characterize the inventory requirements and service level for a hedging policy for a two-stage system subject to a two-state Markov-modulated Poisson demand process. In the fourth section, we formulate an optimization problem to determine the location for the inventory hedge. We report results from our computational experience in the fifth section; these results provide some insight into the behavior and benefit from the hedging policy. In the last section we conclude with a summary and discussion of possible next steps.

Teradyne Case Study

In this section we describe an industrial project with Teradyne Inc.[1], the world's largest supplier of automatic test equipment and software for the electronics and telecommunications industries. The Industrial Consumer Division (ICD) is Teradyne's largest and most profitable finished goods assembly division. ICD manufactures and markets systems that test linear and mixed-signal semiconductor devices. Linear and mixed-signal devices function in a diverse group of commercial products, including personal computer disk drives, stereos, wireless phone systems, VCRs, camcorders, and automobiles.

ICD's Demand

The demand environment for ICD is very volatile. The mean demand rate per week in a down cycle can be on the order of a few testers. However this rate can more than double with no or little forewarning. In Figure 1, we show the aggregate sales for the ICD division for the past five years (the actual revenues have been masked for confidentiality reasons); due to shortages and backorders, the actual demand may have been even more variable. A key reason for this volatility seems to be the presence of a very pronounced bullwhip effect. ICD falls at one of the upstream most positions within its technology supply chains if one were to refer to a computer or a VCR as a true end-product that requires ICD's testers. A typical up cycle can last for one or two quarters, which could be followed by a down cycle of similar length. This volatility causes a great deal of chaos on the ICD supply chain.

[1] Source: Teradyne Web Site, http://www.teradyne.com/

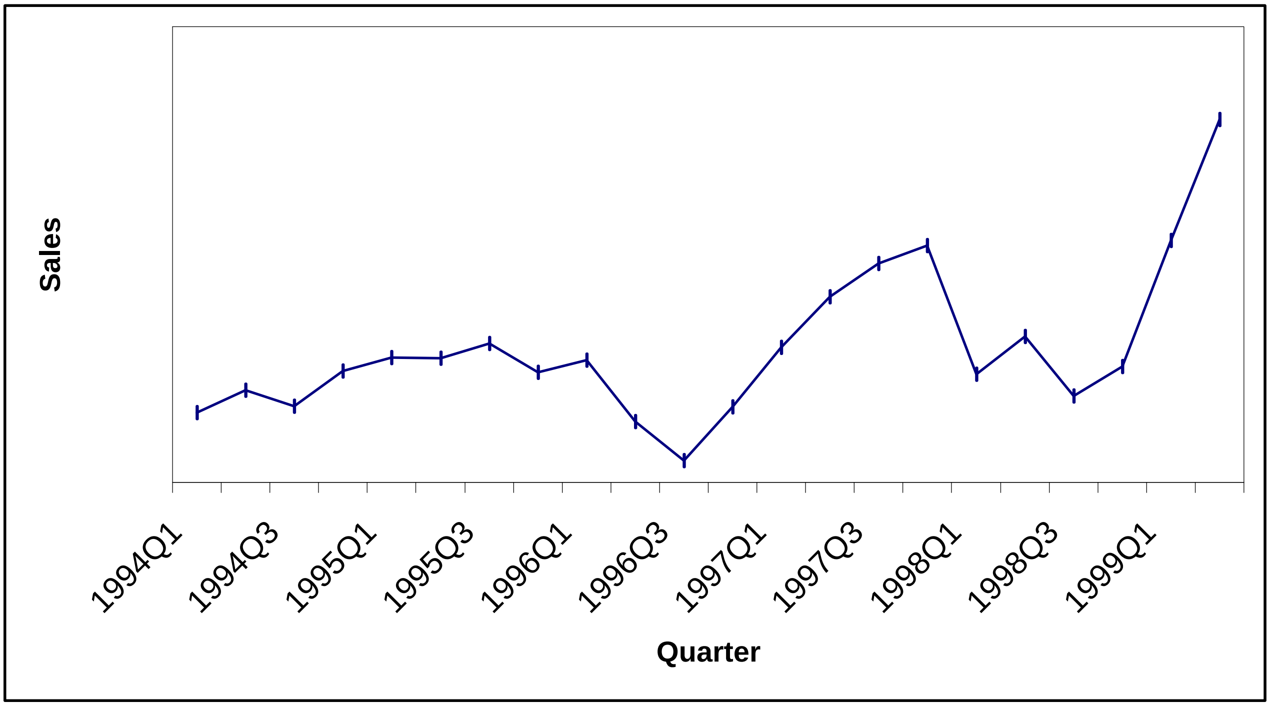

Figure 1: Quarterly Sales Data (Q1 1994 – Q2 1999)

<u>Product Structure</u>

The Catalyst family of products, which were introduced in 1997, represents ICD's flagship product line. These testers sell for an average price of roughly $1.5 million. The Catalyst represents the largest fraction of ICD's revenues.

The bill of material for the Catalyst family has three key levels, an option level, a printed circuit board (PCB) level, and a piece part level. Piece parts are assembled into PCBs that are then tested and assembled into options. A typical option is comprised of from 1 to 8 different PCB's. A tester consists of about 50 different options, which are assembled together with other subassemblies such as a workstation, test head, and mechanical assembly to form a tester. A customer orders a customized tester by specifying a set of options.

The overall product structure for the Catalyst family has an hourglass structure. At the time of the study, there were roughly 10,000 distinct components, several hundred distinct PCBs, and about 175 options. There were on the order of ${}^{175}C_{50}$ possible end-items, as there are about 50 options per tester.

Roughly 150 Catalysts have shipped from the introduction of the product through the first quarter of 1999. Although the number of Catalysts shipped fluctuates from quarter to quarter as seen in Figure 1, the frequency-of-use for most of the individual options is quite stationary. In Figure 2 we show the percent of the total number of distinct options by frequency-of-use. In Figure 3 we show the cumulative cost for an average Catalyst as a function of the options, ranked by frequency of use.

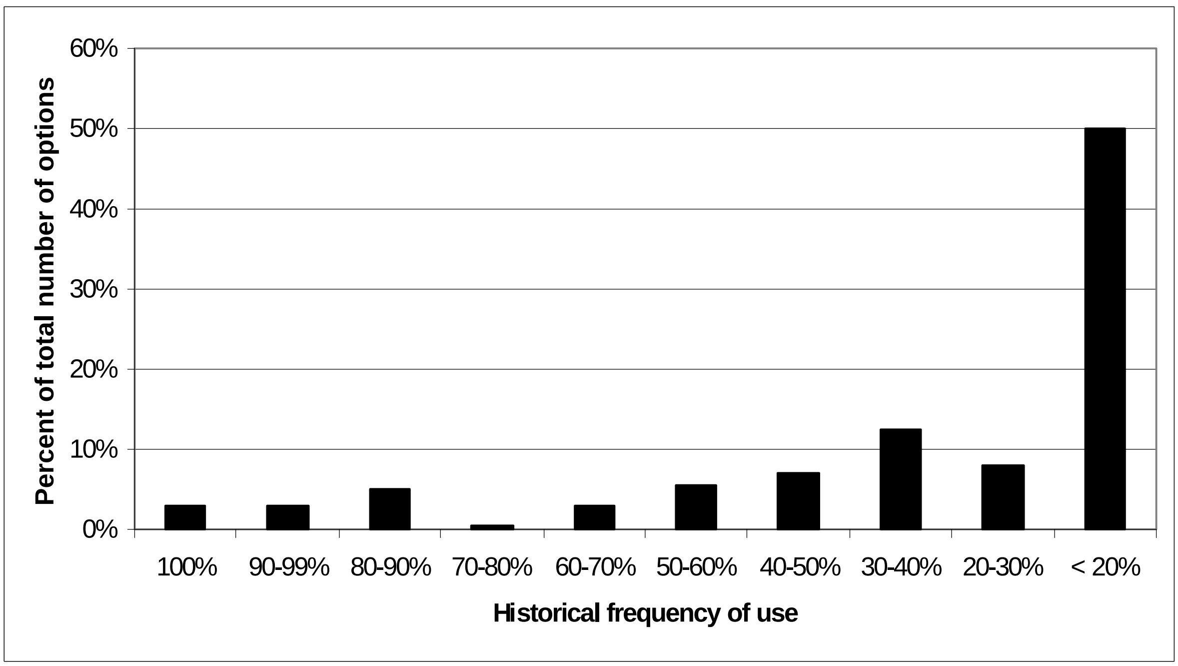

Figure 2: Percent of total number of distinct options versus historical frequency of use

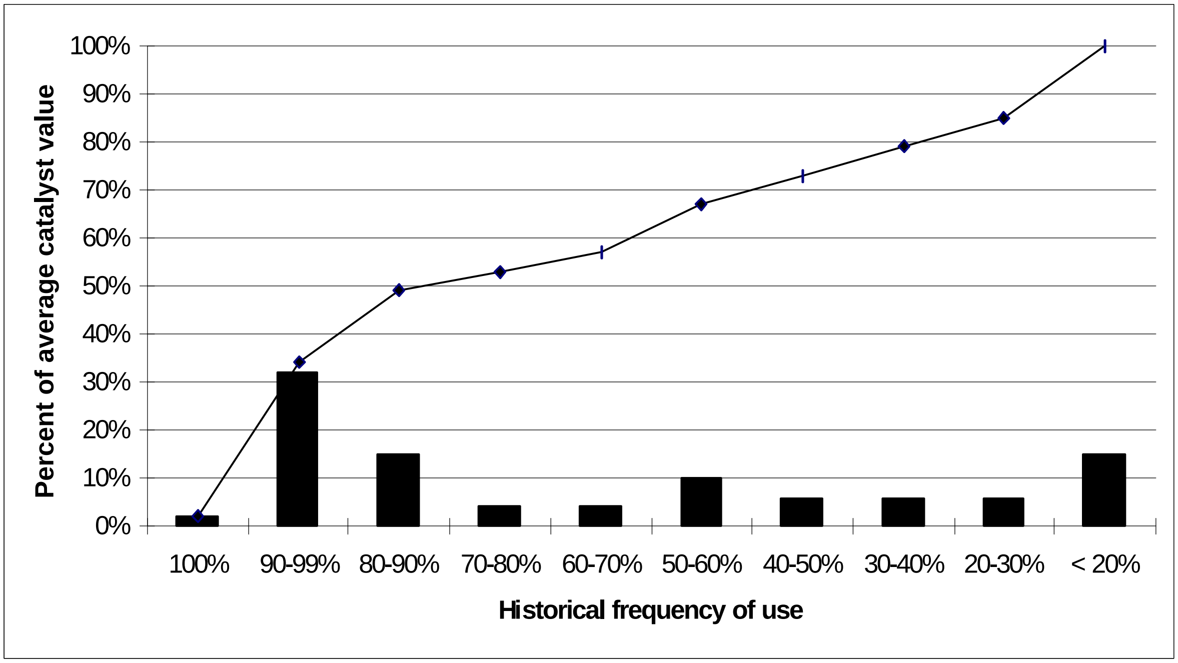

Figure 3: Percent of average catalyst cost versus historical frequency of use

From Figure 2 we note that most Catalysts appear to be quite disparate from one another, as 50% of the options appear in less than 20% of the systems. However, from Figure 3 we see that roughly 72% of the cost of an average system consists of options that are used 40% of the time or more. We also see that the least frequently-used options (those with frequencies-of-use less than or equal to 20%) account for about 15% of the value of an average Catalyst. From these two figures, we see that there is in fact a 80/20 rule in effect, i.e., 10% of the options represent 50% of the costs.

Production Planning

The ICD division operates as a make-to-order division. The cumulative production lead-time (the lead-time for the longest lead-time piece part procured from an outside vendor plus the internal assembly and test lead-times) exceeds the customer lead-time (the delivery lead-time requested by customers). In order to provide competitive customer lead-times, the ICD division must plan and order material and, at times schedule PCB production, prior to receiving an order. The ICD division does this by means of a master production schedule (MPS) that covers a planning horizon of about one year, corresponding to the length of the cumulative production lead-time. The ICD master-scheduling group is responsible for maintaining the MPS.

At the time of the study, the MPS process assumed an aggregate demand rate that would persist for the duration of the planning horizon. That is, for the master schedule planning horizon, ICD planned to sell and to produce x Catalysts per week where x denotes the demand rate. This rate was then revised periodically based on market trends. Thus, if the planning horizon were m weeks, then there would be mx testers in process in the master schedule.

The master scheduling group designates the testers in the MPS into one of three categories: open, identified or booked. An open system is one that has not been allocated to a customer, an identified system is associated with a potential customer, and a booked system is a firm order placed by a customer.

The master-scheduling group plans open testers, assuming a planning bill-of-material, in order to fill the material pipeline. A planning bill-of-material reflects, in theory, the average usage of components and subassemblies. For instance, an option that is used on 70% of the testers would have a planning factor of 0.7, indicating that an average tester requires 0.7 of this option. However, at the time of the study, the planning bill did not reflect the actual option usage, as it had been created prior to the introduction of the Catalyst and had not been revised to reflect the historical usage of options.

Marketing, either through direct contact with potential customers or through market analysis, identifies potential customers for testers. If contact has been made with a customer, marketing requests a tentative product specification for a tester from the customer. Such a tester is henceforth referred to as an identified tester. A tentative due date is also (when possible) obtained from the customer. An open tester from the MPS that falls within the appropriate period of time (closest to the potential due date) is identified and assigned to the customer. Any initial product specifications provided by the customer are substituted for the planning bill. This results in rescheduling to remove unnecessary options and to locate previously unplanned options into the MPS. If no such open tester is available in the MPS, then either a new tester is added to the MPS or the due date is negotiated.

As the tester rolls closer in time within the MPS, the customer either books it or does not commit to the tester, turning it back into an open tester. If the tester books, it typically does not book as initially specified, i.e., the initial product specification changes. This causes additional rescheduling as unnecessary options have to be removed and previously unplanned options have to be located or introduced into the MPS.

In addition, the master schedulers must plan the procurement and production of miscellaneous options that are not part of the planning bill due to the fact that these options were new since the creation of the planning bill. At the time of the study, this planning was done in an ad-hoc manner, based on the experience and judgement of the master-scheduling group. Finally, Teradyne uses the MPS not only to fill the material pipeline with piece-part orders, but also to plan PCB production, by generating from the MPS a time-phased PCB-level requirements schedule for the internal division that assemble boards.

Assessment of the MPS planning process

The planning system has been primarily reactive, as it reacts to problems as they occurred rather than proactively plan for them. Delivery performance to customers has been poor. A great deal of expediting has been required to address material shortages. There has also been constant rescheduling of the MPS, resulting in a great amount of chaos in the entire supply chain, from ICD to the internal suppliers and external vendors.

The MPS process has not had any explicit tactic to accommodate demand volatility. When the demand rate changed, the ability to increase the MPS has been limited by the longest-lead-time parts. If the production lead-time were *m* weeks, then it would take *m* weeks, in theory, for ICD to adapt to a change in its demand rate. In practice, when this happened, the master scheduling group would resort to expediting and rescheduling to try to shorten this response time; in effect, they would incur additional (indirect) costs to attempt to keep up with the change in demand rate. Nevertheless, in spite of these efforts, the overall service performance has been particularly poor when the demand rate changed unexpectedly.

To introduce new open systems into the MPS, the master-scheduling group has used a planning bill created at the time of introduction of the Catalyst family. Planners relied on their experience of option level shortages to plan new options or to make adjustments to the miscellaneous option inventories in the MPS. In this respect the process has been reactive, i.e., options that have run out in the past or have been "hard" to obtain through rescheduling were over buffered, while options that have sporadic use were typically not even planned. Very little attention has been paid to collect and utilize demand histories for each of the options.

Due to the amount of rescheduling at the MPS level together with unanticipated option requirements, the internal suppliers of PCB's have not been able to handle all of the PCB requests from ICD, largely due to not having the right piece-parts. This in turn has delayed the production and ultimate delivery of finished testers. This situation has occurred even in stable demand periods due to the outdated planning bill and customer order changes.

Conceptual Overview of Hedging Policy

The demand uncertainty that ICD faces can be characterized as having three different attributes: time based – when will a customer require a system, option based – what will a customer require, and level based – what is the aggregate demand rate at the tester level. Based on these observations together with a comprehensive study of the demand and cost data for the Catalyst family, we proposed improvements to the current planning process to address these sources of uncertainty.

The major feature of the new planning process is the creation of an intermediate-decoupling inventory as a hedge against aggregate demand uncertainty. In particular, we size this inventory to protect against sudden increases in the aggregate demand rate as happens when a down cycle switches to an up cycle. We also design this inventory so that it protects against uncertainty in the option requirements, and so that it provides relatively short lead times to accommodate customer-specific demand variability.

In terms of the MPS process, we create this intermediate-decoupling inventory by breaking the master schedule into two parts. The first part is the master schedule for the first L weeks, where L is a design parameter for the policy, denoting the hedging point in time. The second part is the remaining master schedule from week L+1 to week m in the future, where m denote the number of weeks in the master schedule

The number of systems in the first part of the MPS should correspond to the maximum reasonable demand over the next L weeks, assuming the current aggregate demand rate and accounting for the current order backlog. The number of systems in the second part of the MPS should correspond to the maximum reasonable demand over an m-L week time period, assuming the maximal aggregate demand rate, i. e., the demand rate during an up cycle.

For instance, suppose m = 16 weeks, L = 4 weeks, and we are currently in a down cycle with demand rate λ_D = 2 systems/week. If the aggregate demand process is Poisson, then we might regard the maximal demand over the L=4 weeks to be, say, 14, corresponding to the 0.98 percentile in the cumulative demand distribution. If the demand rate in an up cycle is λ_H = 4 systems/week, then we might view the maximal demand rate over a 12 week period to be 62, also corresponding to the 0.98 percentile. Then in the master schedule, we would schedule 14 systems in the first 4 weeks, and 62 systems from week 5 to week 16. As before, these systems would be introduced into the master schedule as open testers, then identified with customers and ultimately assigned to orders as time rolls forward.

The second key feature of the hedging policy is with regard to how options are planned. We split the options into two groups, those with fairly stable demand and those without stable demand.

The stable demand group corresponds roughly to those options that are used in more than 20% of the testers, and account for 85% of the material cost of the tester (Figures 2 and 3). These options are included in the new planning bill according to their historical usage rates.

The remaining options are used infrequently and account for less than 15% of the material value of the Catalyst. These options are planned separate from the MPS, in a conservative manner, so as to assure their availability.

Mechanics of Hedging Policy

We depict the operation of the hedging policy in Figure 4. The triangle on the right represents the finished good inventory (FGI) at an option level, the other triangle represents the intermediate-decoupling inventory (INT). The pipeline going into the intermediate-decoupling inventory represents the pipeline inventory between the end of the material pipeline to the INT and the pipeline between the two inventories represents the in-transit material between the INT and the FGI. The circles denote individual systems.

In effect, the policy operates as a two-stage base-stock system, as long as the demand rate remains unchanged. The downstream base stock is set assuming the current demand rate, whereas the upstream base stock is set based on the high demand rate. Each demand triggers a replenishment request from the downstream stage on the upstream stage, where the replenishment time is L weeks. The upstream stage releases a unit from INT, and then will initiate a replenishment of its inventory by scheduling a new tester into its pipeline.

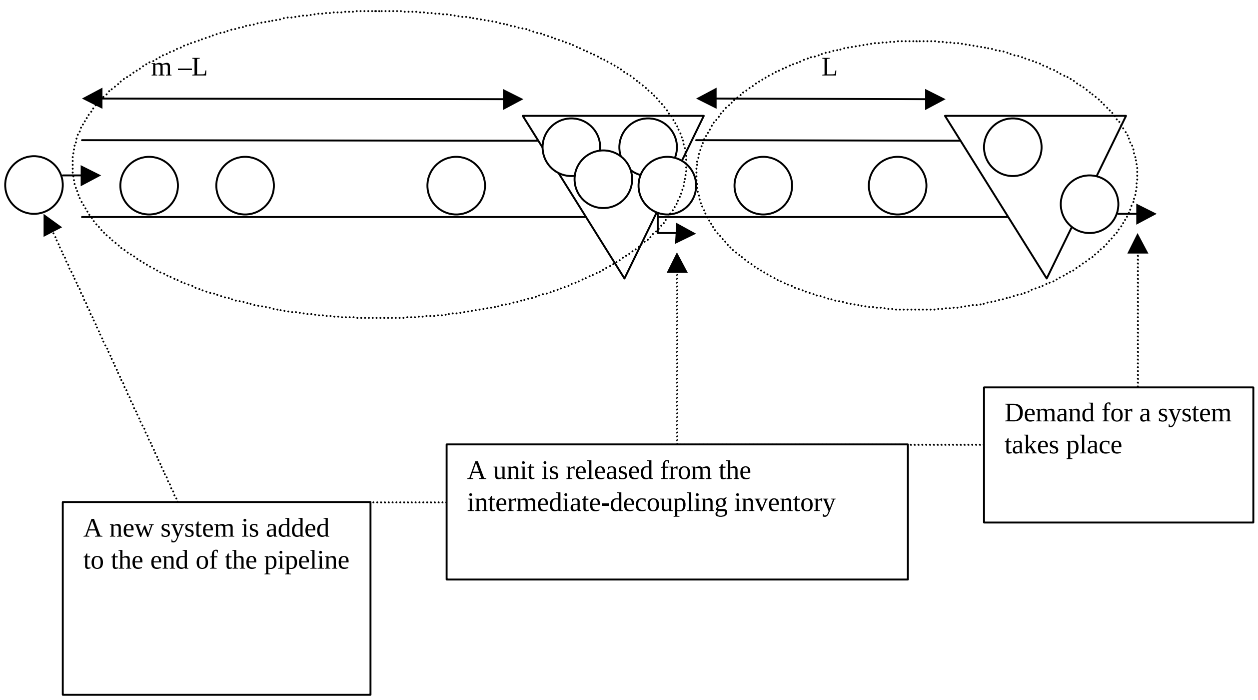

Figure 4: Mechanics of inventory policy. The number of units within each oval corresponds to the base stock.

The contents of each circle are identical, e.g., they represent a planning bill for a tester. As a planning bill moves closer in time, i.e., from left to right in Figure 4, it triggers orders for components with outside vendors. When a planning bill is x time weeks away from FGI, it has triggered orders for all components with lead-times greater than or equal to x weeks. In effect we can think of the planning bill in the pipeline as a pallet, upon which components are "added" as it moves closer to the FGI. Now this "adding" is not literal adding, but rather "virtual" adding as it represent the placement of an order with the vendor for the component.

When the demand rate changes, the policy adjusts the base stock for the downstream stage as quickly as possible. When the demand rate changes from low to high, the FGI base stock must be increased. Once management decides that the rate has changed, they authorize the release of additional material from the INT to ramp the FGI to a new base stock that reflects the new demand rate. Thus, when a demand rate increases, the FGI base stock will be restored to its new target after L time units. If the rate changes from high to low, the FGI base stock must be decreased. Systems will not be released from the INT to the FGI pipeline until the appropriate FGI base stock is reached for the down cycle. In addition, to the extent possible, the excess material is de-expedited and orders are cancelled so as to bring the material pipeline back in line with the new demand rate.

Implementation of Hedging Policy

The implementation of the hedging policy required answers to three questions: Where in time should the intermediate-decoupling inventory be? For sizing the intermediate-decoupling inventory, what demand rate should be assumed for the next up cycle? And what service level target should be used for setting the base stocks for the options?

To determine where to place the hedging point, we examined the shape of the cost accrual profile, which represents the cumulative cost of the components purchased as a function of the cumulative lead-time. In Figure 5, we provide a representative lead-time cost-accrual profile based on the Catalyst. We observed that the long lead-time components represent a small portion of the total material cost of a tester. To this end we picked a point of inflection closest to time t = 0 as a hedging point to put the intermediate-decoupling inventory in order to meet sudden upward spikes in demand.

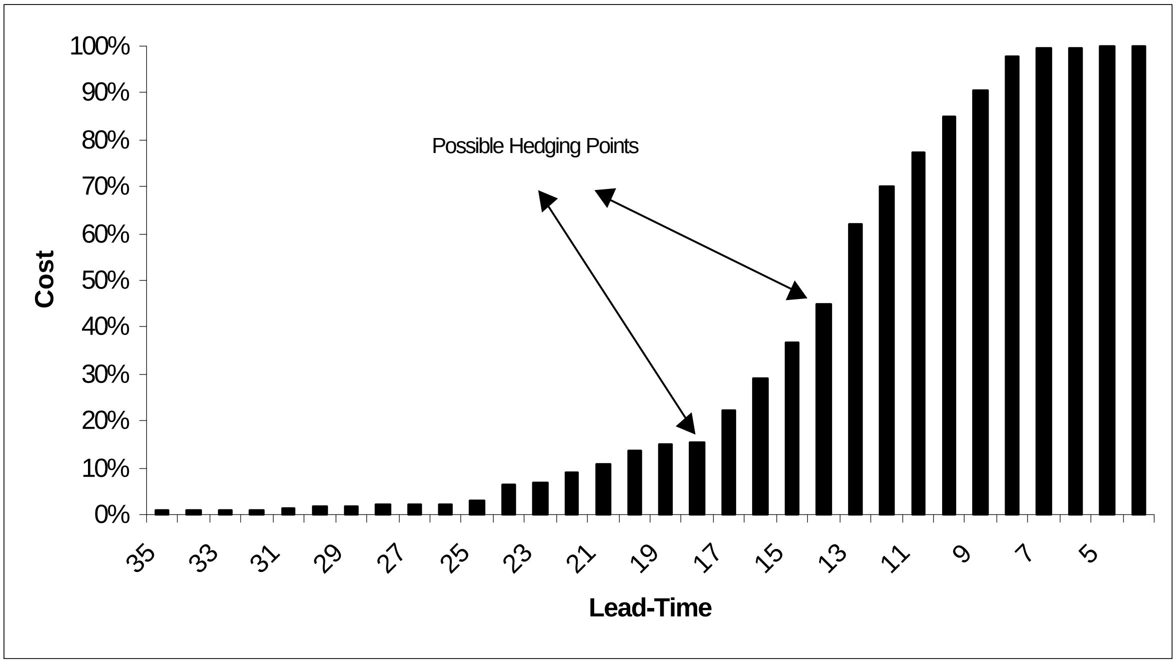

Figure 5: A Sample Lead-Time Cost Accrual Profile (for an average Catalyst)

How much of a ramp should one prepare for? ICD had had no success forecasting demand with traditional forecasting models, due largely to the extreme demand volatility. Thus a different approach which called for expert judgement was undertaken. During this process, a group of Teradyne's senior managers for ICD, with the aid of a decision tree, came to a consensus judgement on the maximum reasonable demand rate for which to plan, after accounting for the impact of not meeting a ramp. ICD's market can be characterized as one with a few very large competitors each trying to capture the other's market-share. The impact of not meeting a ramp can (and has in the past) lead to a significant loss of market share, with serious financial consequences.

The historical fill rate at the option level had been anywhere between 5%-90% depending on the particular option. Due to the product-form nature of end-item versus component fill rates in assembly systems, and due to the implied costs of shortages, we proposed a 97% fill rate as a target fill rate for all relevant options.

Performance Evaluation

Teradyne implemented the hedging policy in the fourth quarter, 1998. The timing of the implementation of the hedging policy could not have been any more opportune. After nearly a year-long slump, a ramp in demand abruptly took place at the start of 1999. The demand rate exceeded the maximal rate that was set by management; however the intermediate-decoupling inventory was sufficient to capture nearly all of the additional demand.

Anecdotally it also seems that the master schedulers have had little difficulty to find the requisite options in the MPS. In the past nearly every order caused some chaos, as at least one option was not available in the required time window within the MPS.

The incremental inventory investment to permit the formation of the intermediate inventory has not been substantial. This is primarily due to the fact that previously, due to the use of an outdated planning bill, the MPS process had resulted in a large amount of inventory in the form of expensive options that had been over planned.

As some indication of the impact of the hedging policy, we cite the following quotation from a "Research Alert" based on upgraded earnings estimates for Teradyne:

"Robust demand for Teradyne's core product, the catalyst mixed signal tester, which is approximately 45 percent of total company sales, is prompting the company to ship more units than it originally forecast."[2]

Also, a press release from Teradyne noted:

"Teradyne, Inc. announced today that Catalyst Test System orders reached nearly 500 units by the end of 1999 due to tremendous demand for digital consumer, communications, and Internet ICs."[3]

Need for Research

In setting up this policy we were unable to characterize the customer service level, even ignoring the assembly nature of the products, during the transition from a low-demand period to a high-demand period. As a consequence we arbitrarily chose a 97% fill rate target for the INT and hoped that this would provide a satisfactory service level over the ramp period. Furthermore we selected the location for the INT by a back of the envelope analysis of the cost accrual profile. In the next section we present a model that is motivated by the hedging policy implemented at Teradyne. With this model, we can gain some quantitative insight into the expected performance of this policy; we can also embed the model in an optimization in order to find the best hedging point.

---

[2] Reuters Securities news release, September 20, 1999, http://biz.yahoo.com/rf/990920/o5.html.
[3] Teradyne press release, February 3, 2000, http://biz.yahoo.com/prnews/0002030ma_teradyn_1.html

Model of Hedging Policy

The intent of this work is to understand better the performance of the hedging policy in contexts, like Teradyne, where demand is very volatile, the lead-times for raw materials are very long, and there is little visibility of the evolution of demand. To do this, we develop a stylized version of the situation faced at Teradyne.

Model framework

We consider a single product with a Markov-modulated Poisson demand process. We assume that demand evolves according to an observable discrete time, recurrent Markov chain with two states, with self-transitions. One state is the high-demand state or up cycle; the other state is the low-demand state or down cycle. In each state, the demand process is Poisson, where λ_H (λ_D) is the demand rate for the high-demand (low-demand) state. The transition matrix for the Markov chain is

$$\begin{pmatrix} 1-p & p \\ q & 1-q \end{pmatrix},$$

where p (q) is the probability of transitioning from the high-demand (low-demand) state to the low-demand (high-demand) state.

We assume that the replenishment lead-time for the product is deterministic and equal to m time units. To simplify the presentation, we assume that the length of the period for the Markov chain equals the replenishment lead-time for the product, namely m time units. For instance, a time unit might represent one week, and both the replenishment lead-time and period length might correspond to a quarter (m=13) or half of a year (m=26). We describe in the last section of the paper how we might relax the assumption of equating the lead-time to the period length.

We assume that there is a finished-goods inventory (FGI) to serve demand, and an intermediate-decoupling inventory (INT) that is L time units away from FGI. That is, the replenishment lead-time for the finished-goods inventory is L time units, provided the intermediate-decoupling inventory is available. The replenishment time for the intermediate-decoupling inventory is m-L time units. We first characterize the performance of the inventory system as a function of the parameter L, and then will formulate an optimization problem in which L is a key decision variable.

There are two pipeline inventories as well: one between the supplier and the INT and the other between the INT and the FGI.

There are no capacity constraints, the external supplier is completely reliable, and excess demand at each of the stages is backlogged.

To relate this stylized model to the Teradyne context, we would view the product as the common planning bill, consisting of a set of components, each with a deterministic replenishment lead-time. The lead-time for the product is the longest of the component lead-times, equal to m time units. Then, the FGI corresponds to a full set of components (or options) necessary to assemble the product. The INT consists of the components with lead times greater than L time units that are managed by means of the master production schedule, as described in the prior section.

<u>The inventory control policy</u>

We assume that each stage operates with a continuous-review base-stock or one-for-one replenishment policy. The FGI follows a state-dependent base-stock policy with a base stock for the high demand (S_H) and a base stock for the low demand (S_D). The INT has a state-independent base stock S_I. When the demand rate changes from low to high, the FGI places a batch order (equal to S_H - S_D) to bring its inventory position up to the high-demand base stock. When the rate changes from high to low, we let the FGI return to the low-demand base stock. Thus, the FGI places no replenishment orders for the next S_H - S_D demands so as to bring its inventory position to the low-demand base stock.

We size the base stock for the INT so that the probability of a stock-out at this stage is suitably small for all demand states. Thus, we must set the base stock to handle the high demand rate as well as the batch order that arises when the demand state transitions from low to high. We assume that any stock-outs that take place at the INT will be handled through expediting, so that they are not a factor in calculating the fill rates of the FGI. This assumption is similar to the "maximum reasonable demand" assumption from Simpson (1958).

Thus, when there is a transition from low to high demand, the FGI places a batch order that will arrive after L time units. Since L is less than the length of a period (m time units), the adjustment from the low-demand base stock to the high-demand base stock is accomplished within a single period, namely the first period of an up cycle.

When there is a transition from high to low demand, we assume that the FGI returns to the low-demand base stock within one period. That is, we assume that with high probability, demand in a low-period exceeds S_H - S_D. This assumption is easy to validate, since the low-period demand is Poisson with mean $m\lambda_D$.

<u>Recurrent cycle</u>

We define a recurrent cycle starting with a transition from a low-demand state to a high-demand state and ending at the next such transition. Thus, a cycle consists of a geometrically-distributed number of high-demand periods, followed by a geometrically-distributed number of low-demand periods. The expected length of a cycle is (1/p + 1/q) periods, or equivalently m(1/p + 1/q) time units.

We decompose this recurrent cycle into four segments or sub-cycles. The first sub-cycle is the single transient period going from low to high demand, in which the FGI will raise its base stock from S_D to S_H. After this transient period, we have the second sub-cycle, a high-demand sub-cycle during which the FGI base stock is S_H; this sub-cycle lasts until there is a transition back to the low demand state. The third sub-cycle occurs when the state returns to low demand, and there is another single transient period while the FGI reduces its base stock from S_H to S_D. This is followed by the fourth sub-cycle, a low-demand sub-cycle during which the FGI base stock is S_D.

<u>Overview of approach and intent</u>

In the following we present a series of claims to characterize the FGI requirements and service level for each sub-cycle. We will then characterize the INT inventory. When the demonstration is straightforward, we state these claims without proof.

The intent is to develop relatively simple closed-form approximations for the inventory and service level for the proposed hedging policy. We can then use these approximations to gain insight into the structure of the policy, to highlight the key tradeoffs, and to frame an optimization for locating the hedging point. We also use the model to illustrate the benefit from creating an inventory hedge, as prescribed by the policy.

To develop the approximations for the inventory and service level, we make three simplifying assumptions:

- To characterize the average inventory level, we treat backorders as negative inventory. This is a common assumption, often justified on the basis that backorders are rare when the service expectations are high.
- To characterize the service level, we assume that demand over the replenishment lead-time has a normal distribution. As we have already assumed that the demand process is Poisson, we note that the Poisson distribution approaches a normal as its mean increases.
- To characterize the inventory in the high to low transient period, we assume that the time for S_H - S_D demands to occur is less than the minimum of (L, m-L). We make this restrictive assumption to simplify the inventory approximation for this sub-cycle. We expect that this assumption is reasonable when $\lambda_H / \lambda_D < 2$; when $\lambda_H / \lambda_D > 2$, we would need to modify our approximation by extending the high to low sub-cycle to multiple periods.

To characterize both the FGI and INT inventories, we use the following standard relationship (e. g., Graves, 1988):

$$X(t) = S - d(t - l, t) \qquad (1)$$

where X(t) denotes the inventory at time t, S is the base stock, *l* is the lead time, and d(s, t) is the demand over the time interval (s, t]. The base stock corresponds to the inventory position, while d(t – *l*, t) is the on-order amount at time t. We will adapt (1) to account for transient periods when we adjust the FGI base stock, as prescribed by the inventory control policy.

To characterize the service level, we use the service factor defined as:

$$z(t) = E[X(t)] \big/ \sigma[X(t)] ,$$

where E[]and σ[] denote the expectation and the standard deviation. Thus, the service factor is the expected inventory, expressed in units of standard deviations.

<u>The low to high transient period</u>

For the first L time units during this period, S_D is the FGI base stock. After L time units, a batch order arrives from the intermediate-decoupling inventory to bring the base stock up to S_H. In the following we develop approximations for the key performance measures for this sub-cycle.

Claim 1: Let t = 0 be the time at which the demand rate transitions from low to high demand. Then, for $0 \le t \le L$, the demand over the time interval (t – L, t] is Poisson with mean $\lambda(t)$, where

$\lambda(t) = (L - t)\lambda_D + t\lambda_H$.

Claim 2: The FGI in the low to high transient period is given by:

$X_{FGI}(t) = S_D - d(t - L, t)$ for $0 \le t < L$

$X_{FGI}(t) = S_H - d(t - L, t)$ for $L \le t \le m$

where d(t – L, t) is Poisson with mean $\lambda(t)$ for $0 \le t < L$, and is Poisson with mean $L\lambda_H$ for $L \le t \le m$.

Claim 3: The average expectation of $X_{FGI}(t)$ in the low to high transient period is:

$$E[X_{FGI}] = \frac{1}{m}\left[S_D L + (S_H - \lambda_H L)(m - L) - (\lambda_D + \lambda_H)\frac{L^2}{2}\right],$$

where we treat backorders as negative inventory.

As noted earlier, we use this to approximate the expected on-hand inventory in the first sub-cycle. This should be a good approximation, as long as the base stocks are set to assure a high level of service. Abhyankar (1999) provides a closed-form refinement of this approximation for the case when $S_D \le L\lambda_H$.

Claim 4: A closed-form approximation for the average safety factor over the first L time units of the low to high transient period is given by:

$$Z = \left(\frac{2}{3}\right)\frac{-(L\lambda_H)^{3/2} + (L\lambda_D)^{3/2} + 3S_D(L\lambda_H)^{1/2} - 3S_D(L\lambda_D)^{1/2}}{(\lambda_H - \lambda_D)L} \qquad (2)\ .$$

Proof: From Claim 2, for $0 \le t < L$, $X_{FGI}(t) = S_D - d(t - L, t)$ where $d(t - L, t)$ is Poisson with mean $\lambda(t)$. Then, we define

$$z(t) = \left(S_D - \lambda(t)\right)\Big/\sqrt{\lambda(t)}\,,$$

to be the safety factor at time t; that is, z(t) denotes the number of standard deviations of protection provided by the base stock S_D as of time t. Thus, we obtain the average safety factor (2) from $Z = \int_0^L z(t)dt / L$. ∵

We propose to use (2) to approximate the fill rate over the first L time units; in effect, we assume that the fill rate is the same as that for normally-distributed demand with Z standard deviations of safety stock.

We tested this approximation across a set of 30 test problems. We set $\lambda_H = 1.2$, $\lambda_D = .7$, and assumed that $S_D = L\lambda_D + z(L\lambda_D)^{1/2}$ for the test problems. We selected z so that the fill rate in the low-demand period was one of four values, namely .9773, .95, .90, .85. We then set L so that the value of S_D fell in the set {6,8,10, …20}. Across this set of test cases this approximation to the fill rate results in an average absolute error of 1.5%.

The safety factor over the time interval (L, m) is $\left(S_H - L\lambda_H\right)\Big/\sqrt{\lambda_H}$, which we can combine with (2) to find the average safety factor for the low to high transient period. We will use this to approximate the fill rate provided by the hedging policy in the low to high transient period.

High-demand sub-cycle

Claim 5: During the high-demand sub-cycle, the FGI and service factor at time t are:

$X_{FGI}(t) = S_H - d(t - L, t)$

$z(t) = (S_H - L\lambda_H)/\sqrt{\lambda_H}$

where d(t – L, t) is Poisson with mean $L\lambda_H$ for all time epochs. Thus, the average inventory (treating backorders as negative inventory) is $S_H - L\lambda_H$, and the average service factor is Z = z(t).

The high to low transient period

When the demand rate switches from high to low, the FGI adjusts its base stock from S_H to S_D. We assume that this occurs by not replenishing the first S_H - S_D demands that occur in the transient period. In the following we characterize the average inventory during this transient period.

Claim 6: The average on-hand FGI during the high to low transient period obtained by treating backorders as negative inventory is:

$$E[X_{FGI}] = \frac{1}{m}\left\{\frac{L^2}{2}(\lambda_H - \lambda_D) + L(S_H - \lambda_H L) + \frac{(S_H - S_D)^2}{2\lambda_D} + (m - L)(S_D - \lambda_D L)\right\}.$$

Proof: Let t = 0 be the time unit at which the demand rate transitions from high to low demand. Define τ to be the time epoch at which the $(S_H - S_D)^{th}$ demand occurs; that is, τ is the first time at which d(0, t) = S_H - S_D.

As in (1), the on-hand inventory is the difference between the inventory position (IP) and the on-order amount (OO). We determine E[IP|τ] and E[OO|τ] to get $E[X_{FGI}|\tau]$; we then find $E[X_{FGI}]$ by taking the expectation over τ.

For a given realization of τ, the expected inventory position is:

$E[IP(t|\tau)] = (t/\tau)(S_H - S_D)$ for $0 \le t \le \tau$

$E[IP(t|\tau)] = S_D$ for $\tau \le t \le m$.

Thus, we find the average inventory position $E[IP|\tau]$ to be:

$$E[IP|\tau] = \frac{1}{m}\left\{\left(\frac{S_H + S_D}{2}\right)\tau + S_D(m - \tau)\right\}.$$

For the on-order amount, we need to distinguish between the orders placed prior to time 0, and the orders placed after time τ. Let $OO_1(t)$ denotes the on-order amount at time t due to orders placed prior to time 0; and let $OO_2(t|\tau)$ denotes the on-order amount at time t, conditioned on τ, due to orders placed after time τ. The former orders are given by:

$OO_1(t) = d(t-L, 0)$ for $0 \le t \le L$

where $d(t - L, 0)$ is Poisson with mean $\lambda_H(L - t)$ for $0 \le t < L$. The orders place after time τ are given by

$OO_2(t|\tau) = d(\tau, t)$ for $\tau < t \le \tau+L$

$OO_2(t|\tau) = d(t-L, t)$ for $\tau+L < t \le m$

where $d(s, t)$ is Poisson with mean $(t-s)\lambda_D$ for $\tau < s < t \le m$. Note that we assume that $\tau+L$ is less than m with probability 1. By taking the expectation at each time t, and averaging over each of these intervals, we obtain the average on-order $E[OO|\tau]$:

$$E[OO|\tau] = \frac{1}{m}\left\{\frac{\lambda_H + \lambda_D}{2}L^2 + L\lambda_D(m - \tau - L)\right\}.$$

We subtract $E[OO|\tau]$ from $E[IP|\tau]$, and take the expectation over τ to get Claim 6. ∵

We do not provide an explicit characterization of the safety factor for this transient period, as the fill rate should not be an issue when the demand rate drops. It is easy to see that the fill rate is no worse than that for the low-demand sub-cycle, given in the next claim.

Low-demand sub-cycle

Claim 7: During the low-demand sub-cycle, the FGI and service factor at time t are:

$X_{FGI}(t) = S_D – d(t – L, t)$

$z(t) = (S_D - L\lambda_D)/\sqrt{\lambda_D}$

where d(t – L, t) is Poisson with mean $L\lambda_D$ for all time epochs. Thus, the average inventory (treating backorders as negative inventory) is $S_D - \lambda_D L$, and the average service factor is Z = z(t).

Intermediate-decoupling inventory

We can characterize the INT inventory using (1), with modifications to account for the transient effects. These claims parallel those for the FGI.

Claim 8: Let t = 0 be the time at which the demand rate transitions from low to high demand. The INT inventory in the low to high transient period is:

$X_{INT}(t) = S_I – d(t – m + L, t) – (S_H - S_D)$ for $0 \leq t < m - L$

$X_{INT}(t) = S_I – d(t – m + L, t)$ for $m - L \leq t \leq m$

where d(t – m + L, t) is Poisson with mean $\lambda_D(m - L - t) + \lambda_H t$ for $0 \leq t < m - L$, and is Poisson with mean $(m-L)\lambda_H$ for $m - L \leq t \leq m$. Thus, the average expectation of $X_{INT}(t)$ in the low to high transient period is:

$$E[X_{INT}] = S_I - \frac{m-L}{m}\left\{(S_H - S_D) + L\lambda_H + \frac{(m-L)}{2}(\lambda_D + \lambda_H)\right\},$$

where we treat backorders as negative inventory.

Claim 9: During the high-demand and low-demand sub-cycles, the INT inventory is:

$X_{INT}(t) = S_I - d(t - m + L, t)$

where $d(t - m + L, t)$ is Poisson with mean $(m-L)\lambda_H$ and $(m-L)\lambda_D$, respectively, for all time epochs. Thus, the average inventory (treating backorders as negative inventory) is $S_I - (m-L)\lambda_H$, and $S_I - (m-L)\lambda_D$, respectively.

Claim 10: The average expectation of $X_{INT}(t)$ in the high to low transient period is:

$$E[X_{INT}] = S_I - \frac{m-L}{m}\left\{(S_D - S_H) + L\lambda_D + \frac{(m-L)}{2}(\lambda_D + \lambda_H)\right\},$$

where we treat backorders as negative inventory.

The development of this approximation parallels that for Claim 6. This development assumes that $\tau < L$ with high probability, where τ is the time epoch at which $d(0, t) = S_H - S_D$.

<u>Average FGI and INT inventories</u>

With these results, we can now determine closed-form approximations for the average FGI and INT by weighting each sub-cycle by its expected duration.

Claim 11: From Claims 3, 5, 6, and 7, the average value of FGI over a cycle, obtained by treating backorders as negative inventory, is:

$$E[X_{FGI}] = \left(\frac{1}{p+q}\right)\left\{p(S_D - L\lambda_D) + q(S_H - L\lambda_H) + pq\frac{(S_H - S_D)^2}{2m\lambda_D}\right\}$$

Claim 12: From Claims 8, 9, and 10, the average value of INT over a cycle, obtained by treating backorders as negative inventory, is:

$$E[X_{INT}] = S_I - \frac{q}{p+q}(m-L)\lambda_H - \frac{p}{p+q}(m-L)\lambda_D .$$

Optimization Model

We now pose an optimization problem to minimize the total expected inventory holding cost, subject to service level constraints.

The *decision variables* for the optimization problem are the location of the intermediate-decoupling inventory, namely the choice of L, and the base stock parameters, S_H, S_D and S_I.

The *objective* is to minimize the total expected inventory holding cost per unit time. As the pipeline inventory does not depend on the decision variables, we state the objective in terms of the inventories FGI and INT:

$$\text{Min } C_{FGI}\, E[X_{FGI}] + C_{INT}\, E[X_{INT}] \qquad (3),$$

where $E[X_{FGI}]$ and $E[X_{INT}]$ are given in Claim 11 and 12, and C_{FGI} and C_{INT} denote the inventory holding costs.

The holding cost for the intermediate-decoupling inventory, C_{INT}, depends on its location, namely on L. Thus,

$$C_{INT} = f(L) \qquad (4),$$

where the function $f(l)$ is the holding cost per unit per time unit for inventory located l time units from FGI. By definition, $C_{FGI} = f(0)$. We expect this function to be non-increasing; as we locate INT further from FGI, the holding costs should decrease. In the context of an assembly product, like for Teradyne, this function corresponds to the holding cost of all material with lead-times exceeding l time units.

We specify *constraints* on the fill rates in the low-demand sub-cycle, in the high-demand sub-cycle and in the low to high transient period. There is no constraint for the high to low transient period, as its fill rate is at least as good as that for the low-demand sub-cycle. We also impose a constraint on INT to assure that it is a decoupling inventory, as we assume in the development of the hedging-policy model. Finally, we constrain L to be non-negative with an upper bound of m.

$$\frac{(S_D - l_D L)}{\sqrt{l_D L}} \geq z_D \tag{5}$$

$$\frac{(S_H - l_H L)}{\sqrt{l_H L}} \geq z_H \tag{6}$$

$$\frac{L}{m}\left(\frac{2}{3}\right)\frac{-(L\,l_H)^{3/2} + (L\,l_D)^{3/2} + 3S_D(L\,l_H)^{1/2} - 3S_D(L\,l_D)^{1/2}}{(l_H - l_D)L} + \frac{(m-L)}{m}\frac{(S_H - l_H L)}{\sqrt{l_H L}} \geq z_{D-H} \tag{7}$$

$$\frac{\left(S_I - (S_H - S_D) - (m-L)\,l_H\right)}{\sqrt{(m-L)\,l_H}} \geq z_I \tag{8}$$

$$0 \leq L \leq m \tag{9}$$

$$S_H, S_D, S_I \geq 0 \tag{10}$$

The first two constraints establish fill rate targets for the low demand and high demand sub-cycles. We assume here that a normal distribution is a good approximation for the Poisson demand over the replenishment lead times. The parameters z_D and z_H represent the desired fill rate targets for these sub-cycles; that is, the constraints assure that the fill rate during the sub-cycle is at least $\Phi(z_D)$ and $\Phi(z_H)$, where $\Phi()$ is the cumulative distribution function for a standard normal random variable. Since we consider a continuous review system with Poisson arrivals and one-for-one replenishment, the stock-out probability is the fill rate.

The third constraint models the average fill rate during the low to high transient period. The first term is the service factor during the first L time units of the low to high transient period (see Claim 4). The second term corresponds to the service factor for the remaining sub-period of length m-L time units, for which the FGI is at the high-demand base stock. The service target during this transient period is particularly critical in a context like at Teradyne; when the demand rate changes, there is a tremendous opportunity to either gain or lose market share depending on the firm's ability to respond.

The fourth constraint establishes the service provided by the intermediate-decoupling inventory. From Claim 8, we see that the maximum demand on this inventory occurs m-L time units into the low to high transient period. At this point of time, the INT needs to have released the batch shipment to bring the FGI base stock up to S_H, and will have been subject to the high rate of demand for m-L time units, namely its lead-time. The above constraint sets a service factor for this time epoch.

The optimization problem, given by (3)-(10), is a relatively small non-linear program: four non-negative decision variables, four service-level constraints, plus an upper bound on L. For all of our numerical tests, we solve the problem using the Excel spreadsheet non-linear optimization solver. As the problem is not a convex program, we have no guarantee that we find globally optimal solutions. Nevertheless, for several test problems, we re-solved the problem from different starting points, and always obtained the same solution. This suggests that the solutions found are possibly globally optimal.

Numerical Experiments

In order to get some intuition about the structure and performance of the hedging policy, we solved a set of test problems. We fixed the demand rates and target safety factors and varied the Markov chain and cost function parameters as follows:

Length of period (time units per period)	$m = 30$
Demand rates (units of demand per time unit)	$\lambda_H = 12.5$, $\lambda_D = 7.5$
Target safety factors	$z_H = z_D = 1.6$, $z_{D\text{-}H} = 1.4$, $z_I = 2.0$
Expected length of up cycle (periods)	$1/p = 1.5, 3, 10$
Expected length of down cycle (periods)	$1/q = 1.5, 3, 10$
Cost shape parameter	$s = 1.36, 2.24, 4.84$

Table 1: Parameters for test problems

The fixed parameters are similar to those from Teradyne. We vary the Markov chain parameters to capture a range of cycle lengths, where short and long cycles have an expected duration of 1.5 periods and 10.0 periods, respectively.

To specify the optimization problem, we also need to determine the function f(*l*), defined as the holding cost per unit per time unit for inventory located *l* time units from FGI. We assumed the following form for f(*l*):

$$f(\ell) = \left(1 - \frac{\ell}{m}\right)^{s}$$

where $s \geq 0$ is a shape parameter. Thus, $C_{FGI} = f(0) = 1$. By varying the shape parameter s, we can model different cost accrual profiles for the supply chain. We set three values for s, given in Table 1, so that $f^{-1}(0.5) = 12$, 8 and 4, respectively. In Figure 6 we plot these three cost functions. Thus, for s=1.36, we might assume that 50% of the cost is incurred in the first 18 time units of the lead-time, and the remaining 50% in the last 12 time units of the lead time. The second and third cost functions are more skewed, with the 50% breakpoint occurring closer to the end of the lead-time. We term these cost functions low, medium and high to reflect the level of skewness.

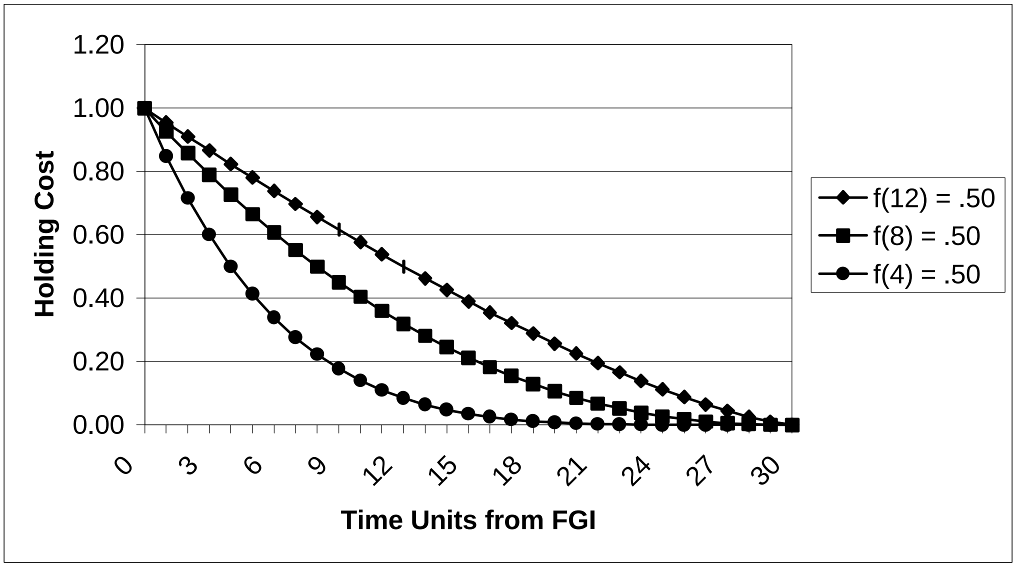

Figure 6: Holding cost functions

In Table 2 we present the results from solving this set of 27 test problems. We report the optimal value for L, and the optimal inventory holding cost. As each test problem has a different average demand rate, we have divided the inventory holding cost by the demand rate to get a holding cost per time unit per unit of demand. We also indicate which constraints are binding in the optimal solution. Finally, for each test problem, we provide the optimal inventory holding cost per unit of demand when there is no intermediate-decoupling inventory. We obtain this result by dropping constraint (8) and adding the constraint L=m to the original optimization problem. The last column in the table gives the percentage savings from the hedging policy, relative to the best policy without an intermediate-decoupling inventory.

1/p	1/q	Cost Shape	Bind. Con-straints	L*	Cost/Unit @ L*	Cost/Unit @ L=30	Savings %
1.5	1.5	low	6, 7, 8	26.9	6.69	6.88	2.7%
1.5	3	low	6, 7, 8	26.2	8.19	8.49	3.5%
1.5	10	low	6, 7, 8	25.6	10.41	10.89	4.4%
3	1.5	low	6, 7, 8	28.3	5.12	5.19	1.3%
3	3	low	6, 7, 8	27.4	6.47	6.61	2.1%
3	10	low	6, 7, 8	26.1	9.14	9.49	3.7%
10	1.5	low	6, 7, 8	29.5	3.44	3.45	0.3%
10	3	low	6, 7, 8	29.2	4.14	4.16	0.5%
10	10	low	6, 7, 8	27.7	6.31	6.42	1.8%
1.5	1.5	med.	6, 7, 8	20.2	5.74	6.88	16.5%
1.5	3	med.	6, 7, 8	20.2	6.99	8.49	17.6%
1.5	10	med.	5, 7, 8	18.4	7.95	10.89	27.0%
3	1.5	med.	6, 7, 8	21.4	4.51	5.19	13.1%
3	3	med.	6, 7, 8	20.6	5.61	6.61	15.2%
3	10	med.	6, 7, 8	19.9	7.81	9.49	17.7%
10	1.5	med.	6, 7, 8	23.7	3.18	3.45	7.7%
10	3	med.	6, 7, 8	22.7	3.76	4.16	9.7%
10	10	med.	6, 7, 8	21.0	5.51	6.42	14.2%
1.5	1.5	high	6, 7, 8	12.8	4.07	6.88	40.9%
1.5	3	high	5, 7, 8	12.4	4.48	8.49	47.3%
1.5	10	high	5, 7, 8	14.8	3.99	10.89	63.4%
3	1.5	high	6, 7, 8	13.2	3.27	5.19	37.0%
3	3	high	6, 7, 8	13.2	4.01	6.61	39.4%
3	10	high	5, 7, 8	13.9	4.12	9.49	56.6%
10	1.5	high	6, 7, 8	13.7	2.44	3.45	29.3%
10	3	high	6, 7, 8	13.7	2.80	4.16	32.6%
10	10	high	6, 7, 8	13.1	3.96	6.42	38.3%

Table 2: Results from test problems

In commenting on these results, we first note that for most problems the binding constraints were (6), (7) and (8). Thus, the fill-rates in the high-demand sub-cycle and in the low to high transient period equal the targets, whereas the fill rate in the low-demand sub-cycle exceeds the target. The service constraint (6) for the high-demand sub-cycle determines the high-demand base stock S_H, while the service constraint (7) for the low to high transient period determines the low-demand base stock S_D. Thus, the policy results in excess safety stock during the down cycle in order to satisfy the service requirement during the ramp up from the low demand to high demand state.

For three of the test problems, though, the binding constraints were (5), (7) and (8). Here the fill rate in the high-demand sub-cycle exceeds the target, and the high-demand base stock S_H is set to satisfy the service constraint (7) for the low to high transient period. For these instances, it seems that it is cheaper to have excess safety stock in the up cycle, rather than in the down cycle, due to the length of the down cycle.

With this understanding of the structure of the solutions, we make two observations from this set of test problems.

The benefit from the hedging policy and the location of INT are very sensitive to the shape of the holding cost function. The cost savings relative to the no-hedging policy grow as the cost function becomes more skewed. Similarly, we find that the location of INT, given by L, moves closer to FGI, as the cost function becomes more skewed.

As explanation, we note that the location of the INT influences the FGI in a couple of ways. First, the location of INT determines the replenishment lead-time for the FGI; the closer INT is to FGI, the shorter is the replenishment time for FGI, thus reducing the size of the state-dependent FGI base stocks. Second, the location of INT is critical to determining the service level for the low to high transient period. For smaller values of L, the system transitions more quickly from the low-demand state to the high-demand state, and the service level in the low to high transient period depends more on S_H, rather than on S_D, as is clear from (7). Thus, the closer

INT is to FGI, the easier it is to satisfy the fill-rate constraint for the low to high transient period, thus reducing the size of the state-dependent FGI base stocks.

For these two reasons, we reduce the FGI base stocks as we locate the INT closer to FGI. With a more skewed holding cost function, the cost for creating the INT is less. Thus, we locate the INT closer to the FGI and get greater relative cost savings from the hedging policy, as the holding cost function gets more skewed.

In light of this observation, we note that reducing the lead-time of a component increases the skew of the holding cost function. Focusing on the most expensive components yields the biggest change to the shape of the cost function. Thus, the implementation of a hedging policy makes these lead-time reduction efforts even more valuable.

For a given holding cost function, the location of INT is relatively insensitive to the length of the up and down cycle, while the benefit from the hedging policy increases as the length of the down cycle increases relative to the up cycle.

To explain this observation, we first note that the Markov chain parameters, p and q, do not appear in the constraints (4)-(10), but only impact the objective function through the expressions for $E[X_{FGI}]$ and $E[X_{INT}]$ (see Claims 11 and 12). As we increase (decrease) the length of the down (up) cycle, we increase the weight in the objective function on the expected inventory during the low-demand state.

From Table 2, we see that for each holding cost function the choice for L is fairly stable over the range of values for the Markov chain parameters, p and q. Indeed, as we increase (decrease) the length of the down (up) cycle, the change to L depends on whether constraint (5) or (6) is binding. If constraint (6) is binding, then there is excess safety stock during the down cycle and L decreases slightly so as to reduce the low-demand FGI base stock, as determined by the fill-rate constraint (7). If constraint (5) is binding, then there is excess safety stock during the up cycle and L increases slightly so as to reduce the high-demand FGI base stock, as determined by the fill-rate constraint (7). Nevertheless, the overall sensitivity of L to the cycle lengths seems modest for these test problems.

However, the benefit of the hedging policy increases as the length of the down cycle increases. The no-hedging policy must maintain excess FGI safety stock during the down cycle in order to satisfy the fill-rate constraint during the low to high transient period. Thus, the advantage of the hedging policy over the no-hedging policy grows as we increase (decrease) the length of the down (up) cycle, as can be seen in Table 2.

To understand better the role of the intermediate-decoupling inventory, we conducted a second set of experiments in which we fix the location of the inventory and solve for the optimal base stocks. We set the expected length of the up cycle $1/p = 3$, and assumed a holding cost function with high skew ($s = 4.84$). In Figure 7, we plot the cost per unit of demand as a function of L; we do this for three choices for the expected length of the down cycle ($1/q = 1.5$, 3, 10). From this figure we see that the optimal choice of L is quite insensitive to the length of the down cycle. However, the shape of the cost function does depend on the length of the down cycle. As the length increases, the cost penalty for deviating from the optimal choice of L grows. Furthermore, we see that it is better to err on the side of a larger-than-optimal choice for L, rather than less than the optimum.

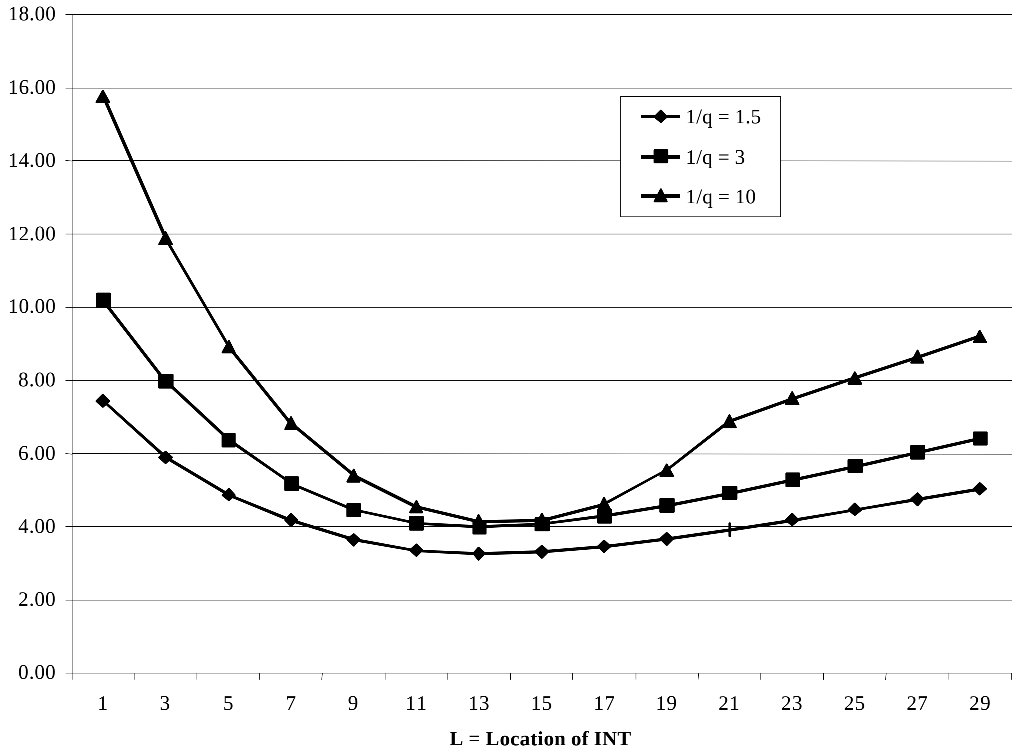

Figure 7: Optimal cost per unit of demand, for fixed L, for 1/p = 3, s = 4.84

Conclusion

In this paper we have studied a hedging policy for protecting a supply chain against cyclic variability. This work is motivated by a successful application at Tereadyne, where we helped to implement an inventory hedge. To understand this tactic better, we develop a simple model of a two-stage supply chain, subject to non-stationary demand in the form of a two-state Markov modulated Poisson process. For this system we develop closed-form approximations for the inventory and for the customer service level. We can then embed these approximations into an optimization model to highlight the trade-offs between inventory investment and customer service. This optimization finds the optimal location of the inventory hedge, and permits exploration of the sensitivity of the hedging policy to the various system parameters.

A key result is that the hedging policy can be a very effective tactic for handling extreme demand uncertainty in the form of cyclic variability. The benefits from the hedging policy grow as the holding cost for the supply chain becomes more skewed. The hedging policy benefits also increase as the length of the down cycle, relative to the up cycle, increases.

The analysis relies on several simplifying assumptions that merit further discussion. In particular, we assume that the length of the period equals (or exceeds) the replenishment lead-time for the product. We also assume that the time for S_H - S_D demands to occur is less than the minimum of (L, m-L). These assumptions were made so that the adjustment of the FGI base stock occurs within a single period, both when the demand rate changes from low to high and when it changes back from high to low. If we were to relax these assumptions, then it is possible to extend the analysis by considering a multiple-period transient to accomplish the base-stock adjustment when there is a state change. Needless to say, though, the required analysis is much more convoluted, and we doubt that it would yield any new insights.

We conclude with some unanswered questions. The development of the hedging policy provides a means to create some flexibility in the supply chain, namely the ability to ramp up to the high demand rate in L time units. But this assumes that there are no relevant capacity constraints for components with lead-times less than L. An interesting question is how to incorporate capacity limitations in the supply base, of one form or another, into this planning tactic.

The analysis also assumes that there is no forewarning of the change in the demand rate, but that the firm recognizes the change immediately once it happens. Of course reality is not so clean. There is often some advance warning or forecast of major shifts in demand. An interesting question is how to adapt the hedging policy so as to use this information.

References

Abhyankar, H. S. (2000), Inventory Control for High Technology Capital Equipment Firms, Ph.D. Thesis, A. P. Sloan School of Management, M.I.T., Cambridge MA, 134 pp.

Baker K. R., (1993) "Requirements Planning," in Handbooks in Operations Research and Management Science Vol. 4, Logistics of Production and Inventory, edited by S. C. Graves, A. H. G. Rinnooy Kan and P. H. Zipkin, North Holland, Amsterdam.

Chen, F. and J.-S. Song, (1997), "Optimal Policies for Multi-Echelon Inventory Problems with Nonstationary Demand," working paper, 25 pp.

Graves, S. C., (1988), "Safety Stocks in Manufacturing Systems," *Journal of Manufacturing and Operations Management*, 1, 67-101.

Guerrero, H. H., K. R. Baker and M. H. Southard, (1986), "The Dynamics of Hedging the Master *International Journal of Production Research*, 24, 1475-1483.

Miller, J. G., (1979), "Hedging the Master Schedule," in Disaggregation, Problems in Manufacturing and Service Organizations, edited by L. P. Ritzman, L.J. Krajewski, W. L. Berry, S. H. Goodman, S. T. Hardy, and L. D. Vitt, Martinus Nijhoff Publishing, Boston.

Simpson, K. F., (1958), "In-Process Inventories," *Operations Research*, 6, 863-873.

Song, J.-S., and P.H. Zipkin, (1992), "Evaluation of Base-Stock Policies in Multiechelon Inventory Systems with State-Dependent Demands. Part I: State-Independent Policies", *Naval Research Logistics*, 39, pp. 715-728.

Song, J.-S., and P.H. Zipkin, (1996), "Evaluation of Base-Stock Policies in Multiechelon Inventory Systems with State-Dependent Demands. Part II: State-Dependent Depot Policies", *Naval Research Logistics*, 43, pp. 381-396.

Vollmann, T. E., W. L. Berry, and D. C. Whybark, (1992), Manufacturing Planning and Control Systems, 3rd Edition, Richard D. Irwin, Inc. Homewood Illinois.

Wijngaard, J. and J. C. Wortman, (1985), "MRP and Inventories," *European Journal of Operational Research*, 20, 281-293.