

**AN EMPIRICAL STUDY OF SERVICE DIFFERENTIATION  
FOR WEAPON SYSTEM SERVICE PARTS**

**Vinayak Deshpande**

Krannert School of Management, Purdue University  
W. Lafayette, IN 47907  
e-mail: vinayak@purdue.edu

**Morris A. Cohen**

Department of Operations and Information Management  
The Wharton School of the University of Pennsylvania  
Philadelphia, PA 19104-6366  
e-mail: cohen@wharton.upenn.edu

**Karen Donohue**

Operations and Management Science Department  
Carlson School of Management of the University of Minnesota  
Minneapolis, MN 55455  
email: kdonohue@csom.umn.edu

August 2000  
revised February, 2002, August 2002

*Subject classifications:* Military, logistics: weapon systems. Inventory/Production,  
multi-item: service differentiation.

## Abstract

The question of how to effectively manage items with heterogeneous attributes and differing service requirements has become increasingly important to supply chains that support the delivery of after sales service. However, there has been little investigation to date on how organizations actually manage inventory levels under such circumstances. This study provides such an investigation, focusing on the logistic system used to manage consumable service parts for weapon systems in the U.S. military. Our findings, based on interviews and rigorous analysis of part attribute and performance data, suggest that in practice a part's service level is negatively affected by an item's cost and is less affected by attributes such as its priority code. We introduce a simple inventory model to explain our empirical findings and explore how variations in item attributes can interact with an inventory policy to affect system performance. Based on this model, we recommend using explicit service level targets for priority categories to achieve performance consistent with part priority. We show, using military data, that a service differentiation strategy can be an effective way of allocating inventory investment by providing higher service for critical parts at the expense of accepting lower service levels for parts with less importance.

## 1 Introduction

The delivery of differentiated levels of service to disparate classes of customers is an increasingly important requirement in today's "customer centric" environment. In this context, the fundamental challenge for supply chain managers is to support the need for customized service while still exploiting the cost savings inherent in scale economies and risk pooling opportunities within their supply chain networks. A case in point is the U.S. military, which recently moved the management of consumable service parts for weapon systems from the individual services (i.e., Navy, Army, Air Force, and Marines) to the central Defense Logistics Agency (DLA). This new arrangement offers an opportunity to allocate inventory expenditure more efficiently by taking advantage of economies of scale in ordering and inventory pooling. However, developing a system to accommodate such a wide range of parts for a diverse customer base is a difficult undertaking. The complexity of real environments makes it difficult to gauge how policy characteristics, such as budget constraints, service targets, and priority codes, will play off each other.

Our goals in this paper are (1) to provide insight into the conflicts that can arise when balancing service differentiation and cost containment in a real, complex, logistics environment and (2) to define and test ways to reduce these conflicts. We do this through a series of data analysis and descriptive modeling drawn from an empirical study of the U.S. military’s service parts logistics system. While our analysis draws from the military environment, the lessons learned apply to any company struggling to manage product inventories with heterogeneous attributes and diverse customer service requirements.

We begin by analyzing how part attributes affect actual performance (i.e., fill rate and response time) in practice. In particular, we perform a statistical study to test the relationship between part attributes and performance under DLA’s current logistics system. The study highlights the role of *criticality* and *essentiality* codes used by the military to indicate a service part’s importance to the mission. The objective behind this classification is to provide tailored response time and service targets for parts within each priority category. However, we find that after controlling for other part attributes, a part’s essentiality and criticality are not significant drivers of performance. Instead, cost appears to be the dominant performance driver. Our study appears to be the first to quantify this management problem, which may arise in any complex inventory system that serves multiple customer classes under severe budget constraints.

Based on interaction with DLA, we provide a simple inventory model to explain our empirical findings. Our model shows that if aggregate service level constraints are used in inventory investment allocation decisions, major emphasis may be placed on certain part attributes, such as cost, while other attributes, such as criticality and essentiality may be de-emphasized. We offer some simple remedies for dealing with this problem, including

developing explicit service targets for different categories of parts. We illustrate the potential benefits of this idea with an inventory model, using data drawn from DLA. The results suggest that differentiated service requirements can be met without a significant increase in inventory costs for reasonable ranges of service differentiation.

These findings provide important managerial insights for companies trying to consolidate their inventory management across different product divisions, customer groups or part categories. A key challenge in trying to implement such a policy lies in the ability of the logistics system to provide differentiated service across customer segments in a cost effective fashion. An important conclusion of our study is that explicit service level targets should be used for the different priority categories, so as to align service level with a part's importance. Service differentiation strategies can be an effective way of utilizing inventory investment because they provide higher service for the more important parts at the expense of accepting lower service levels for parts with less impact. Category specific service targets also provide a means for controlling for variation in other attributes (such as cost) for any given part population. Finally, this study suggests new opportunities for analytical research in rationing and incentive issues in service differentiation, as highlighted in the conclusion section and considered in our follow-on research.

As in previous empirical studies of inventory systems (e.g., Cohen and Lee 1988, Cohen and Zhang 1997, and [Lee and Billington 1993](#)), our research focuses on describing what is currently being done and then offers lessons for other companies based on these observations. Our observations are statistically grounded rather than anecdotal since we had access to a large database of part and performance measures. Our findings confirm that service differentiated classes of service parts can be used to meet customer needs while containing

costs. This idea builds on the work of Cohen and Lee (1990) who studied two service part management systems; one used by a computer firm and the other by an automobile manufacturer. They highlight the necessity to group parts based on their “essentiality” and to use a different service requirement for each category.

There has been a wide stream of literature on service parts logistics dating back thirty years. Sherbrooke (1968) developed the well-known METRIC (Multi-Echelon Technique for Recoverable Item Control) model for management of repairable items. This seminal work generated a whole new research area in multi-echelon inventory control as illustrated by the works of Simon (1971), Deuermeier and Schwarz (1981), Graves (1985), [Cohen et al. \(1986, 1988, 1989, and 1990\)](#), Lee (1987), Cohen, Kleindorfer and Lee (1988) and Wang, [Cohen and Zheng \(1999, 2000, and 2001\)](#). A recent paper by [Rustenburg et al. \(2000\)](#) analyzes spares parts management for technical systems under budget constraints, similar to the environment studied here. The benefits of using these techniques are enormous, as illustrated by [Muckstadt and Thomas \(1980\)](#) and [Cohen et al. \(1990\)](#). However, most of the analytical research on service part management systems focuses on determining optimal inventory levels for given inventory holding costs and target service rates. Most of this literature is silent on how target service levels are actually set in practice. Our study sheds light on this issue by explicitly considering the impact of part attributes on observed service measures in a real service parts management system.

In the next section we describe the current military service parts logistics system in more detail. Section 3 then outlines and tests a series of hypotheses concerning the drivers of performance for this system. In section 4, we offer a descriptive inventory model whose drivers are consistent with our empirical results. We use this model to illustrate how explicit

service level constraints for different demand classes can help to effectively allocate inventory in light of a budget constraint. Finally, in section 5, we outline what lessons our study offers other organizations who are contemplating consolidating their inventory management services.

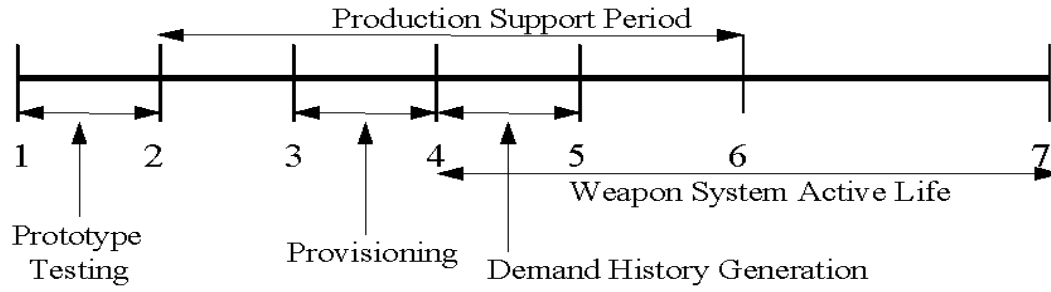
## 2 Process Description

The U.S. military’s logistics system stocks and services hundreds of thousands of consumable service parts to support a worldwide installed base of weapon systems. The overall investment in these parts exceeds \$10 billion. While our study focuses on consumable parts, it is worth noting that weapon systems are supported by a combination of repairable and consumable service parts. The systems used to support repairable parts are quite complex and, for the most part, still reside within the individual military services. The logistics system for consumable parts, in contrast, is now centrally controlled through the Defense Logistics Agency (DLA).

### 2.1 Demand Lifecycle

The demand for consumable parts comes directly from customers on ships, aircraft, or other military equipment. This demand is driven both by initial deployments to support system rollout as well as requests for replacement parts generated by failures in the field. Weapon systems typically are assemblies of both consumable and repairable items at various levels of indenture. Examples include a 5” 54 mounted gun and a CIWS (i.e., Close-in-Weapon-Systems, which are automated firing systems used on ships to hit close range targets). Bombs and rockets are not weapon systems, but rather munitions.

Figure 1 illustrates the life cycle of a typical weapon system. The decision to introduce



1. Decision made to introduce weapon system
2. Production of system starts
3. Provisioning of system begins
4. System introduction begins
5. Early introduction ends
6. Production support ends
7. System retired

Figure 1: Weapon System Life Cycle

a weapon system is made at time point 1. Planning for the support of the weapon system begins at this stage. After about one to two years, the production of the system and its parts begin. At time point 3, the provisioning for the system begins and inventory targets are tentatively set throughout the supply chain. At time point 4, which is about three to five years from time point 1, weapon system introduction begins. The active life of a weapon system is typically fifteen to twenty years, with production support ending about three to five years before the system is retired. Thus there is a demand ramp-up stage when the system is being introduced, followed by a stable active life stage, and finally a phase out stage between time points 6 and 7. In our analysis, we focus on DLA's stocking decisions for consumable service parts during the active life of a weapon system (i.e., between time points 4 and 6).

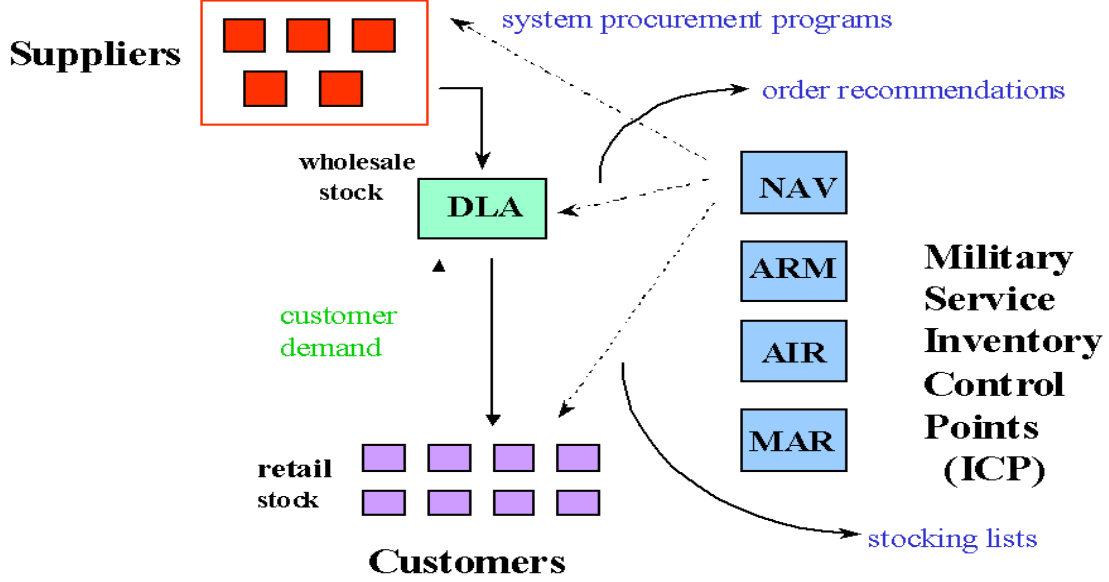


Figure 2: Service Parts Logistics System

## 2.2 Supply Chain Structure

Figure 2 provides an overview of the supply chain structure for consumable parts supporting active weapon systems (see Cohen et al. 1998 for more details). DLA manages order requests directly from end customers within each military service. However, each military service also has a central advisory group within its organization which serves as an information intermediary between DLA and its customers within that service. We refer to these advisory groups as ICPs since each functions within the respective service’s Inventory Control Point (ICP). Each ICP is responsible for providing DLA with demand forecasts and order recommendations for their respective client base.

Most consumable service parts are stocked at two levels, referred to as retail and wholesale. Recommended inventory targets at both levels are typically determined by procedures based on Sherbrooke’s METRIC model. Retail inventories are located at the final customer sites (i.e., on ships and at bases in the field). Once retail inventory targets are established



for these sites, replenishments are managed according to a basestock policy (i.e., orders are placed to maintain the target inventory level).

Wholesale inventories, which support multiple systems for multiple retail customers at multiple stocking locations, are also typically replenished using a base stock policy. However, determining appropriate stocking targets at the wholesale level requires more complex procedures to consolidate demand and service requirements across systems, customers and locations while meeting aggregate budget constraints. DLA currently allocates inventory investments at the wholesale level to balance two conflicting goals: minimizing holding and investment costs while maximizing mission readiness. This trade-off is made more complex by the fact that parts differ in how important they are in supporting readiness. A part's importance is measured by two dimensions: 1) its *essentiality* to the operation of the weapon system in which it is housed, and 2) the *criticality* of that system to the customer's overall mission.

A part is assigned one of four essentiality Codes depending on whether failure of the part renders the weapon system inoperable (very high), affects personnel safety (high), degrades the operational effectiveness of the weapon system (but does not render inoperable) (medium), or does not affect the operation of the weapon system (low). The part is then assigned a criticality Code based on the importance of its system application (high, medium, or low). Together these codes dictate the Weapon System Indicator Code (WSIC) for a part, see Table 1. These codes are assigned by each ICP and passed on to DLA with the implicit assumption that DLA will provide better service (i.e., higher fill rates and shorter response time) for parts with higher WSIC values.

| <b>Part Essentiality Code</b>   | <b>Very High (VH)</b> | <b>High (H)</b> | <b>Medium (M)</b> | <b>Low (L)</b> |
|---------------------------------|-----------------------|-----------------|-------------------|----------------|
| <b>Mission Criticality Code</b> |                       |                 |                   |                |
| <b>High (H)</b>                 | A                     | B               | C                 | D              |
| <b>Medium (M)</b>               | E                     | F               | G                 | H              |
| <b>Low (L)</b>                  | I                     | J               | K                 | L              |

Table 1: Weapon System Indicator Codes (A-L)

DLA develops its own policy for setting stocking targets, and for allocating inventory investments, across parts at the wholesale level. The entire process for wholesale stock level planning is iterative, beginning with recommendations received by DLA from the four military services via their ICPs. These order recommendations are based on METRIC model results and managerial input, influenced by each part’s estimated failure rate, service specific WSIC classification, system population, costs and other factors. DLA aggregates these order recommendations across the four services and modifies them based on its own budget, service, and minimum order quantity constraints. These decisions are complicated by the fact that different ICP’s may have different essentiality and criticality classifications for a common part. In such cases DLA uses a “round-up” policy in which DLA assigns the highest WSIC code reported for a part by the multiple users. This issue and its managerial implications are analyzed, in depth, in a companion paper ([Deshpande et al. 2001](#)). DLA’s primary service criteria is an aggregate fill rate target, currently set (arbitrarily) at 90%. The final wholesale stocking targets reflect a compromise between the ICPs and DLA, and are often lower than that suggested by the ICPs.

Because of the complex interaction between the ICPs and DLA, it is difficult to predict the impact of part characteristics (such as cost, essentiality, criticality, etc.) on performance. During our interviews with members from the Navy and the DLA, for example, we discovered

that members differed in their perception about which factors actually or should affect performance based on their individual experience. Since DLA operates under an aggregate fill rate target and overall budget constraint, it is possible that they are not adequately incorporating criticality and essentiality measures in their wholesale stocking decisions. It is important to point out that during the period of our study, the Department of Defense (DOD) was in the midst of downsizing. For DLA, this meant that the annual appropriation for parts replenishment was based on a fraction of the previous year’s consumption. This drives total inventory investment down and as a result the current budget allocation would be based on what they were unable to sell in the previous budget period. This made investments in slow moving, expensive inventory particularly unattractive. On the other hand, since all ICPs focus on their own service priorities and do not face the overall budget constraint imposed on DLA, it is possible that their order recommendations put too much emphasis on criticality and essentiality. Given these conflicting views, it is unclear how cost, criticality, and essentiality will be accounted for in the final wholesale level stocking problem.

In the next section, we compare part attribute information with actual performance data at the wholesale level to analyze the relationship between attributes and performance derived from DLA’s stocking decisions. We use a statistical regression model to rigorously analyze the data in an unbiased fashion to identify which part characteristics, if any, actually affect performance.

### **3 Empirical Findings: The Drivers of Performance**

Service parts performance within DLA is measured in two ways: fill rate and logistics response time (as defined in Table 2).

| Performance Measures    | Variable Name | Description  |
|-------------------------|---------------|--|
| Fill Rate               | FR            | The percentage of customer requests filled from stock, for a given part. This was computed as a summary statistic using transaction data (for data sets 2 and 3).  |
| Logistics Response Time | LRT           | The time needed to fulfill a customer request for a part, from the time of the initial request to the time of receipt. This was captured both as a yearly average (in data set 1) and by transaction (in data sets 2 and 3). |
| ICP Processing Time     | ISPT          | Time taken by ICP to process a transaction (a component of LRT).   |

Table 2: Part-Specific Performance Measures

Based on our interviews with representatives from DLA and the Navy ICP, we identified ten potential drivers of these performance dimensions (see Table 3). In this section, we test the significance of these candidate drivers. We focus special attention on two candidates, part essentiality and weapon criticality, since they are viewed as the most relevant service level criteria in the eyes of the end customer (based on input from Navy ICP).

### 3.1 Data Collection and Preliminary Analysis

Figure 3 gives an overview of the data sets collected for this study and the timing of our analysis. Our initial study examined part-specific attributes for twenty-eight weapon systems. These data are referred to as **Data Set 1** in Figure 3. We collected data on all associated service parts managed by DLA, resulting in a total of 280,000 parts. Time-dependent part attributes (e.g., demand frequency, average demand quantity) reflected annual part activity between the first quarters of 1995 and 1996. In addition to the candidate drivers

| Part Attribute           | Variable Name | Description   |
|--------------------------|---------------|---|
| Standard Unit Price      | SUP           | The price DLA charges its customers, which is a fixed percentage markup over cost.  |
| Essentiality Code        | ES            | The essentiality of the part to the performance of the weapon system. Four categories are possible, as illustrated in Table 1.        |
| Weapon Criticality Code  | WC            | The criticality of the weapon system to the success of the mission. Three values are possible, as illustrated in Table 1.             |
| Administrative Lead Time | ALT           | The expected/planned administrative processing time required for processing replenishment orders. (A constant time for a given part). |
| Production Lead Time     | PLT           | The expected/planned lead time faced by DLA for its replenishment orders from its suppliers. (A constant time for a given part).      |
| Part Commonality         | PC            | The number of weapon systems containing the part.   |
| Annual Demand            | AD            | The total number of units requested (across all customer requests) in a year.   |
| Annual Demand Frequency  | DF            | The number of customer requests for a part in a year (where the quantity of the order may vary widely by request).                    |
| Acquisition Code J       | AJ            | Implies the part is non-stocked and centrally acquired.   |
| Acquisition Code Z       | AZ            | Implies that the part is an insurance item*.  |

\* Items that are generally not subject to periodic replacement but subject to replacement as the result of accidents or other unexpected occurrences.

Table 3: Candidate Performance Drivers

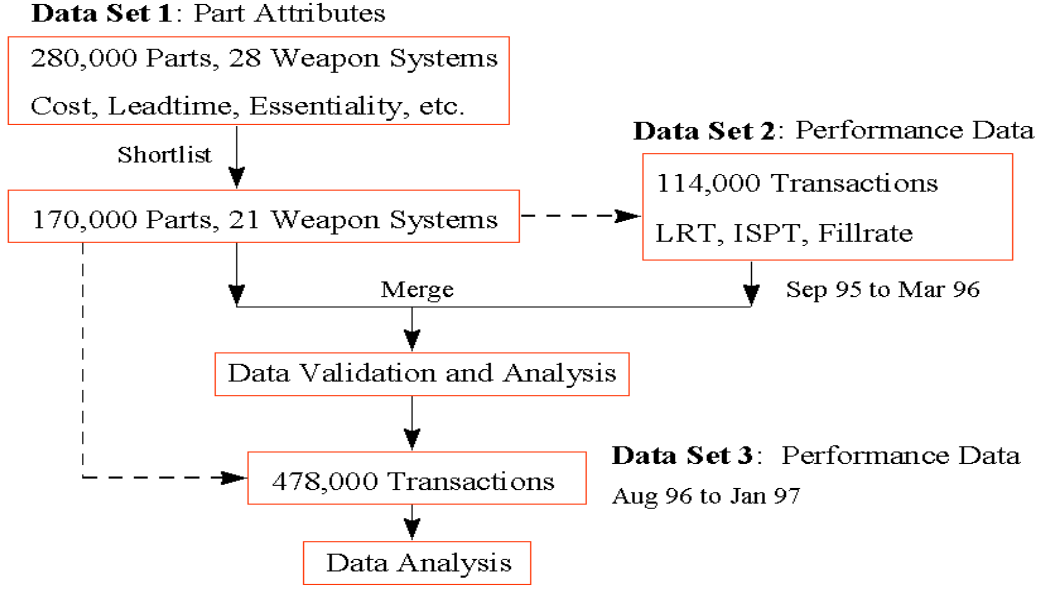


Figure 3: Flow Chart for Data Analysis

outlined in Table 3, we also collected an average logistics response time (LRT) measure for each part that served as a summary performance measure for this time period. The purpose of this initial study was to characterize the general population of service parts for naval weapon systems and help identify a representative subset of weapon systems to use for the transaction-specific data set.

This population of service parts contained a wide range of demand frequencies and standard unit prices. Many parts (42%) displayed no demand activity over the twelve-month sample. The majority of parts (70%) had demand frequency of four or fewer in this period. Demand frequency averaged 12 with a standard deviation of 72. Annual demand quantity (total number of items requested for all issues) was also low, averaging 613 units with 64% of parts displaying demand of 10 units or fewer. We observed that high demand parts tended to be less expensive. Administrative lead times varied from zero days to almost one year, with 73% having a lead time of 100 or fewer days. Production lead times are generally longer

than administrative lead times, with some as long as two years. The complicated production process for complex weapon systems, combined with the thorough supplier qualification and contract award process, often leads to such large lead times.

Based on this preliminary analysis, we selected a representative subset of twenty-one weapon systems from the original set of twenty-six. Our criteria for selection, based on input from the Navy ICP, included:

1. Maximize the diversity of weapon system attributes, such as Mission Criticality, Life Cycle (age), Population Density, Complexity (number of parts) and System Type (i.e., Ships, Surface Warfare, and Aviation).
2. Maintain a similar distribution of part attributes as seen in the larger weapon system population.
3. Reduce the number of parts by at least 30% subject to criteria 1 and 2, limiting the total number of parts to fewer than 200,000. (Note: Nav ICP recommended this specific criteria as a way to control the size of our transaction-based data collection. They felt that collecting and transmitting transaction-based data for more than 200,000 parts would be extremely difficult given the one to many mapping between a part and its transactions.)

In order to satisfy criteria 1 and 2, we omitted five weapon systems. This resulted in a 38 percent reduction in parts, satisfying criteria 3. Table 4 illustrates how the two data sets vary with respect to five key part-specific attributes. The two sets are nearly identical in terms of cost and average production and administrative lead times. The reduced set displays a slightly higher average demand frequency as well as a higher proportion of essential parts (i.e.,

parts having 'Very High' essentiality, see Table 1 for terminology). Since we are particularly interested in the performance of critical parts, having a higher proportion of such parts was not seen as an issue. The five removed weapon systems were, for the most part, either new or no longer supported. Parts for these systems, therefore, exhibited a lower demand frequency, on average.

|                  | Production<br>Lead Time<br>Mean, Std | Admin.<br>Lead Time<br>Mean, Std | Demand<br>Frequency<br>Mean, Std | Standard<br>Unit Price<br>Mean, Std | Essent.<br>% of 'VH'<br>Parts |
|------------------|--------------------------------------|----------------------------------|----------------------------------|-------------------------------------|-------------------------------|
| Original<br>data | 174, 100                             | 91, 51                           | 12, 72                           | 227.2, 256                          | 26%                           |
| Reduced<br>Data  | 173, 102                             | 92, 51                           | 16, 86                           | 242.5, 314                          | 33%                           |

The original data set of 26 weapon systems consisted of 276,000 unique parts.

The reduced data set of 21 Weapon Systems consists of 170,959 unique parts, a reduction of 38%.

Table 4: Comparison of Part Attributes

Next, we collected transaction data for parts corresponding to this reduced list of twenty-one weapon systems. These transaction-based data are referred to as **Data Set 2** in Figure 3. We tracked a total of 113,805 transactions over a period of five months, starting with the first quarter of 1996. Of the 170,959 parts contained in the reduced part set only 31,000 (18%) displayed demand activity over this period. As mentioned earlier, the key performance measures collected include Fill Rate (FR) and Logistics Response Time (LRT), see Table 2.

After analyzing this data set, the project team concluded that a larger data set, covering a more current time period, was required. Additional transaction data consisting of 478,921 transactions over the period August 1996 to January 1997 was then collected. These data



| Standard<br>Unit Price (\$) | Very Low (VL)<br>[\$0,\$150] | Low (L)<br>[\$150,\$500] | Medium (M)<br>[\$500,\$1000] | High (H)<br>[\$10000, \$200000] | Very High (VH)<br>[\$200000+] |
|-----------------------------|------------------------------|--------------------------|------------------------------|---------------------------------|-------------------------------|
| Fill Rate                   | 95.7%                        | 93.1%                    | 89.2%                        | 75.7%                           | 54.6%                         |
| Average LRT                 | 26.3 days                    | 29.0 days                | 33.7 days                    | 54.8 days                       | 97.4 days                     |
| No. of<br>Transactions      | 156650<br>(32.71%)           | 86372<br>(18.03%)        | 168436<br>(35.17%)           | 63321<br>(13.22%)               | 4142<br>(0.87%)               |
| No. of<br>Parts             | 33838<br>(19.8%)             | 25122<br>(14.7%)         | 69726<br>(40.8%)             | 38794<br>(22.7%)                | 3417<br>(2.0%)                |

Table 5: Average Response Time (LRT) by Cost Categories

are referred to as **Data Set 3** in Figure 3. See Cohen et al. (1998) for a detailed analysis of the data sets including different breakdowns and charts.

The transaction data exhibited a wide range of performance characteristics. The overall fill rate across all parts was 90.6%, quite close to the stated goal of 90%. The mean LRT for an immediate fill (i.e., orders filled from stock) was 24 days, whereas the mean LRT for a delayed fill was 123 days. Table 5 illustrates how fill rate and average LRT vary with part cost. We developed these cost categories during our interviews with representatives from Navy and DLA, based on how the military looks at cost breakdowns while making decisions. The striking negative relationship between part cost and service performance, (statistically significant at  $p < 0.01$ ), raises some interesting questions. Is this pattern appropriate or necessary? Are the underlying causes due to explicit managerial policy or are they a result of environmental and structural factors beyond the control of system managers?

Table 6 illustrates the relationship between performance and the WSIC codes. One would expect service to be highest for parts having high essentiality which reside in systems having high criticality (i.e., parts in the upper left corner of Table 6). Our data confirms this intuition. However, one would also expect service to drop off as one moves to the right or down

| System<br>Criticality | Part Essentiality |                |                |                |
|-----------------------|-------------------|----------------|----------------|----------------|
|                       | Very High (VH)    | High (H)       | Medium (M)     | Low (L)        |
| <b>HIGH (H)</b>       |                   |                |                |                |
| Fill Rate             | 92.8%             | 70.5%          | 84.5%          | 79.5%          |
| Average LRT (days)    | 29.9              | 54.3           | 37.4           | 46.5           |
| No. of Transactions   | 320,455 (70.05%)  | 14,267 (3.12%) | 18,892 (4.13%) | 27,536 (6.02%) |
| No. of Parts          | 35971 (47.9%)     | 2617 (3.5%)    | 5390 (7.2%)    | 7603 (10.1%)   |
| <b>MEDIUM (M)</b>     |                   |                |                |                |
| Fill Rate             | 90.8%             | 77.4%          | 82.4%          | 76.6%          |
| Average LRT (days)    | 33.86             | 50.18          | 47.0           | 55.74          |
| No. of Transactions   | 38,786 (8.48%)    | 3777 (0.83%)   | 4228 (0.92%)   | 8949 (1.96%)   |
| No. of Parts          | 9346 (12.4%)      | 2887 (3.8%)    | 1041 (1.4%)    | 2140 (2.8%)    |
| <b>LOW (L)</b>        |                   |                |                |                |
| Fill Rate             | 88.1%             | 90.4%          | 80.1%          | 73.0%          |
| Average LRT (days)    | 38.8              | 34.1           | 45.2           | 54.4           |
| No. of Transactions   | 17,013 (3.72%)    | 407 (0.09%)    | 1584 (0.35%)   | 1558 (0.34%)   |
| No. of Parts          | 6191 (8.2%)       | 177 (0.2%)     | 555 (0.7%)     | 1196 (1.6%)    |

Table 6: Average Logistic Response Time (LRT) by Criticality/Essentiality Codes

this matrix. This trend holds true for all essentiality categories except (H). For parts in this category, service appears to improve as system criticality decreases. This raises the question, what is different about essentiality (H) items? It also raises the more general question, are these performance level differences really attributed to a part’s essentiality/criticality rating or are they driven by differences in other characteristics of parts within these categories? For example, do some essentiality/criticality categories have a higher proportion of low cost parts (which are easier to service)?

In order to develop a better understanding of the complex performance/attribute relationship we computed correlations across key attributes (see Table 7). Most correlations reported in this table are close to zero and all of them are less than 0.3. The most significant correlations are between essentiality or criticality with demand frequency. It is not surprising that parts classified most important are used more frequently. It is also interesting to

note that both essentiality and weapon criticality are negatively correlated with cost (i.e., standard unit price) which, as noted earlier, is strongly negatively correlated with service. The cost/criticality correlation is much stronger than the cost/essentiality correlation, however. This difference helps to explain the results observed in Table 6. Parts that are critical tend to be less expensive and hence will exhibit higher service. The variation of cost with essentiality is much weaker. Review of the data indicated that the average price for the four essentiality codes (VH, H, M and L) are \$96, \$279, \$208, and \$240 respectively. Thus, essentiality code H parts are the most expensive and thus we would expect them to have lower service (as they do in Table 6).

|                          | Essentiality | Weapon Criticality | Standard Unit Price | Production lead-time | Administrative lead-time | Demand Frequency |
|--------------------------|--------------|--------------------|---------------------|----------------------|--------------------------|------------------|
| Essentiality             | 1.000        | -0.050***          | -0.060***           | -0.080***            | -0.080***                | 0.230***         |
| Weapon Criticality       | -0.050**     | 1.000              | -0.150***           | -0.050***            | -0.050***                | 0.260***         |
| Standard Unit Price      | -0.060***    | -0.150***          | 1.000               | 0.063***             | 0.033***                 | -0.010***        |
| Production lead-time     | -0.080***    | -0.050***          | 0.063***            | 1.000                | 0.280***                 | -0.060***        |
| Administrative lead-time | -0.080***    | -0.050***          | 0.033***            | 0.280***             | 1.000                    | -0.020***        |
| Demand Frequency         | 0.230***     | 0.260***           | -0.010***           | -0.060***            | -0.020***                | 1.000            |

Note: All correlation coefficients are Pearson correlation coefficients

except for WSIC correlations where we report Spearman Ranked Correlations

\*\*\* - Significant at  $p < 0.0001$  level.

Table 7: Correlation table

We test the significance of these relationships more rigorously in the next subsection through a regression analysis. We use regression analysis to sort out the underlying significance of each attribute in explaining the observed variation in service performance, given the

complex interactions that exist across the various attribute variables. As we shall see, the regression analysis can be used to definitively test hypotheses concerning the relationship of key attribute variables (i.e., cost, essentiality and criticality) and performance.

### 3.2 Focus on Cost and Priority

Our interview data from the Navy ICP revealed early on that customers expect a part to be serviced according to its criticality and essentiality level, irrespective of the part's cost. The data in Tables 5 and 6 suggest this may not be the case. Our purpose here is to test the relationship between these three attributes (cost, criticality, essentiality) and performance in a more rigorous fashion, controlling for all other part attributes. Specifically, we set out to test the following hypotheses.

**Hypothesis 1** *Part cost does not affect performance.*

**Hypothesis 2** *Mission Criticality Code does not affect performance.*

**Hypothesis 3** *Part Essentiality Code does not affect performance.*

The customers' view of how DLA's system should operate, given its mission to support system readiness in a cost-effective manner, is that Hypothesis 1 is true and Hypotheses 2 and 3 are false.

To test these hypotheses, we regressed Logistic Response Time (LRT), against the candidate performance drivers listed in Table 3. Similar regressions were also performed using Fill Rate as the dependent variable. Results from those regressions are similar and therefore not reported here. We introduced a series of indicator variables to depict our ordinal attributes, Part Essentiality (ES) and Weapon Criticality (WC), as well as our categorization of Standard Unit Price (SUP) as listed in Table 6. These variables are set to 1 if the

part lies in that particular category and 0 otherwise. These variables include  $SUP_i, i = vh$  (very high),  $h$  (high),  $m$  (medium),  $l$  (low);  $WC_j, j = h$  (high),  $m$  (medium);  $ES_k, k = vh$  (very high),  $h$  (high),  $m$  (medium). The first (low) category is the default category, and not included as an indicator variable. We also used second order interactions between these indicator variables to test the significance of all possible pairwise combinations of attributes in order to isolate the primary effect of each individual attribute variable. The final regression is then

$$\begin{aligned}
LRT = & \alpha + \sum_{i=costgroups} \beta_{1,i} SUP_i + \sum_{j=crit.groups} \beta_{2,j} WC_j + \sum_{k=ESgroups} \beta_{3,k} ES_k + \beta_4 ALT \\
& + \beta_5 PLT + \beta_6 PC + \beta_7 AD + \beta_8 DF + \beta_9 AJ + \beta_{10} AZ + \sum_{i=costgroups} \sum_{k=ESgroups} \beta_{11,i,k} \\
& + \sum_{k=ESgroups} \sum_{j=crit.groups} \beta_{12,k,j} + \sum_{i=costgroups} \sum_{j=crit.groups} \beta_{13,i,j} \\
& + \text{second order terms.}
\end{aligned} \tag{1}$$

In keeping with customers expectations, if Hypothesis 1 is true than the  $\beta_{1i}$ 's in our regression should be insignificant. On the other hand, if Hypotheses 2 and 3 are false, the  $\beta_{2j}$ 's and  $\beta_{3k}$ 's should be significant.

We summarize the results in Table 8 for both a normal ordinary least squares (OLS) regression and a rank regression. The OLS regression uses the absolute value of LRT for the regression, while the ranked regression uses the percentile score of the LRT as the dependent variable. Thus the rank regression results in a rescaling of the dependent variable, which can result in a change in the estimated slopes for the two regressions. Our hypothesis tests are based on the statistical significance of the independent variables and not on the magnitude of the coefficients in the regression. As long as the independent variable shows statistical significance in both the regressions, we conclude that the two models agree. These

| Coefficient Estimates | Variable Name        | OLS Regression | Rank Regression |
|-----------------------|----------------------|----------------|-----------------|
| $\alpha$              | Intercept            | 21.770***      | 47.320***       |
| $\beta_{1,vh}$        | $SUP_{vh}$           | 56.350***      | 23.260***       |
| $\beta_{1,h}$         | $SUP_h$              | 22.250***      | 14.430***       |
| $\beta_{1,m}$         | $SUP_m$              | 9.330***       | 8.030***        |
| $\beta_{1,l}$         | $SUP_l$              | -1.040*        | -0.590*         |
| $\beta_{2,vh}$        | $ES_{vh}$            | 3.220          | 5.302           |
| $\beta_{2,h}$         | $ES_h$               | 0.100          | -4.710          |
| $\beta_{2,m}$         | $ES_m$               | -2.500         | -1.420          |
| $\beta_{3,h}$         | $WC_h$               | 0.150          | -0.480          |
| $\beta_{3,m}$         | $WC_m$               | -2.250         | -0.350**        |
| $\beta_4$             | $ALT$                | 0.023***       | -0.006***       |
| $\beta_5$             | $PLT$                | 0.017***       | -0.005***       |
| $\beta_6$             | $PC$                 | -0.038***      | -0.017***       |
| $\beta_7$             | $AD$                 | 0.000*         | 0.000*          |
| $\beta_8$             | $DF$                 | 0.007***       | 0.005***        |
| $\beta_9$             | $AJ$                 | 66.810***      | 21.040***       |
| $\beta_{10}$          | $AZ$                 | 4.130***       | -0.891***       |
| $\beta_{11,l,vh}$     | $SUP_l * ES_{vh}$    | 5.680          | 2.310           |
| $\beta_{11,l,h}$      | $SUP_l * ES_h$       | -7.920         | -10.230**       |
| $\beta_{11,l,m}$      | $SUP_l * ES_m$       | 4.010          | 1.400           |
| $\beta_{11,m,vh}$     | $SUP_m * ES_{vh}$    | -4.230         | -6.250          |
| $\beta_{11,m,h}$      | $SUP_m * ES_h$       | 16.030*        | 5.620           |
| $\beta_{11,m,m}$      | $SUP_m * ES_m$       | 1.060*         | -2.920*         |
| $\beta_{11,h,vh}$     | $SUP_h * ES_{vh}$    | 5.120*         | 4.590*          |
| $\beta_{11,h,h}$      | $SUP_h * ES_h$       | -14.460        | -10.580**       |
| $\beta_{11,h,m}$      | $SUP_h * ES_m$       | 6.030          | 1.810           |
| $\beta_{11,vh,vh}$    | $SUP_{vh} * ES_{vh}$ | 7.160          | 10.230*         |
| $\beta_{11,vh,h}$     | $SUP_{vh} * ES_h$    | -34.950        | -25.650         |
| $\beta_{11,vh,m}$     | $SUP_{vh} * ES_m$    | 5.210          | 4.540           |
| $\beta_{11,vh,h}$     | $ES_{vh} * WC_h$     | 2.600          | -1.350          |
| $\beta_{12,vh,m}$     | $ES_{vh} * WC_m$     | 23.90*         | 0.860           |
| $\beta_{12,h,h}$      | $ES_h * WC_h$        | 8.780*         | 10.350*         |
| $\beta_{12,h,m}$      | $ES_h * WC_m$        | 21.240*        | 1.860           |
| $\beta_{12,m,h}$      | $ES_m * WB_h$        | -3.350         | -3.030**        |
| $\beta_{12,m,m}$      | $ES_m * WC_m$        | 2.010          | 3.850**         |
| $\beta_{13,l,h}$      | $SUP_l * WC_h$       | 5.310*         | 1.680           |
| $\beta_{13,l,m}$      | $SUP_l * WC_m$       | 6.750**        | 3.510**         |
| $\beta_{13,m,h}$      | $SUP_m * WC_h$       | -3.350         | -8.160**        |
| $\beta_{13,m,m}$      | $SUP_m * WC_m$       | 2.010          | -1.720          |
| $\beta_{13,h,h}$      | $SUP_h * WB_h$       | 8.780***       | 8.150***        |
| $\beta_{13,h,m}$      | $SUP_h * WC_m$       | 21.240***      | 1.140**         |
| $\beta_{13,vh,h}$     | $SUP_{vh} * WB_h$    | 2.60**         | 12.850*         |
| $\beta_{13,vh,m}$     | $SUP_{vh} * WC_m$    | 23.90***       | 3.460**         |
| Observations          | N                    | 457,552        | 457,552         |
| F Statistics          |                      | 374.940***     | 165.250***      |
| Adjusted $R^2$        |                      | 0.095          | 0.043           |

\*Significant at  $p < 0.1$  level

\*\*Significant at  $p < 0.05$  level

\*\*\*Significant at  $p < 0.0001$  level

Table 8: Results of regression analysis for Logistics Response Time

two separate regressions were performed to control for any functional relationship between response time and the independent variables. The adjusted  $R^2$  value for the OLS and rank regressions was 0.095 and 0.043, respectively, with the F statistics for both models significant at  $p < 0.0001$ . Note that given the complex nature of the military environment, it is difficult to model a precise relationship between the response time and part attribute variables, resulting in a low  $R^2$  value. We therefore used a large sample size to identify these effects with sufficient precision. The F statistics for the model and most variables is significant to a high degree ( $p < 0.0001$ ). We used various methods such as Ordinary Least Squares Regression, Rank Regression, distribution plots etc., and in each case observed similar results regarding the impact of standard unit price, criticality and essentiality codes on performance. Hence, we conclude that our results are robust to any functional relationship between response time and the independent variables. See Hitt and Frei 1999, for a similar justification.

The results indicate that Standard Unit Price is significant at the  $p < 0.0001$  level, refuting Hypothesis 1. To further investigate the direction of this relationship, we conducted Tukey-Kramer and Scheffe tests for pairwise comparisons of means between the cost groups. All pairwise comparisons were significant at  $p < 0.0001$  level, refuting the hypothesis  $\beta_{1,vh} \leq \beta_{1,h} \leq \beta_{1,m} \leq \beta_{1,l}$ . This suggests that, all else being equal, response time increases with cost categories from ‘Low’ to ‘Very High’.

On the other hand, the coefficients for essentiality and criticality codes are not generally significant, providing support for Hypotheses 2 and 3. Only essentiality code ‘M’ showed any significance ( $p < 0.05$ ) and only within the rank regression model. We also were not able to detect a significant directional relationship using the Tukey-Kramer and Scheffe tests. In particular, we could not reject the hypotheses  $\beta_{2,vh} \leq \beta_{2,h} \leq \beta_{2,m}$ , and  $\beta_{3,h} \leq \beta_{3,m}$ .

The fact that essentiality and criticality do not exhibit a significant positive relationship with response time may seem, at first, contradictory with the trend observed in Table 6. However, what the regression has done is shown that once one controls for other part attribute differences (especially cost), the impact of priority codes seems to disappear. We did not see this subtle point when aggregating parts within priority categories in Table 6. The regression suggests that the partial consistency of performance with respect to essentiality and criticality codes seen in Table 6 is due to the part composition within criticality and essentiality categories rather than the category itself.

It is worth noting that most of our other control variables, including production and administrative lead-times, part commonality, demand frequency, and acquisition codes were also significant in the regression model. Acquisition code J, in particular, exhibits a large positive relationship with response time. Recall that acquisition code J items are, by definition, not-stocked and hence should experience longer response times when demand finally comes in for these items. Finally, a large number of interaction terms came out to be insignificant in our regression model. We did not observe any conclusive trend for these terms.

## **4 Managing Service Categories: Descriptive and Prescriptive Models**

Our analysis of Hypotheses 1, 2, and 3 suggests a disconnect between the expectations of the Navy ICP and the actual performance provided by DLA. However, it is premature to claim that DLA completely ignores priority codes and bases its inventory stocking decisions solely on cost. Here we introduce a descriptive model to show how stocking decisions would be made if cost were the primary driver. While this model is quite simple, it captures the basic



tradeoffs DLA considers when setting inventory levels across product groups. After comparing this model with DLA's observed behavior, we introduce a normative model designed to better align the objectives of DLA and its customers.

## 4.1 A Descriptive Model

We conjecture that DLA sets its stocking levels with the objective of minimizing its aggregate inventory investment across all parts, subject to meeting an aggregate fill rate constraint. The following model tests this conjecture by showing how our sample service parts would be supported under this objective. We make the following model assumptions to simplify our analysis while keeping true to DLA's environment. First, we assume demand for each part  $i$  follows a Poisson process with an annual demand rate of  $\lambda_i$  units per year ( $\lambda = \sum_{i=1}^N \lambda_i$ ), which is a reasonable assumption given the low rate of demand for these service parts. We also assume a  $(Q, r)$  replenishment policy for all parts (i.e., DLA places an order of size  $Q$  for a part whenever its inventory position falls below  $r$ ) with a fixed replenishment lead-time  $\tau$ . DLA's major costs include a holding cost, incurred at a rate  $h$  for all inventory held on-hand, and a setup/transaction cost, charged as  $k$  for each replenishment order placed.

DLA's stocking problem is then,

$$\min \sum_{i=1}^N h c_i (r_i - \mu_i + \frac{Q_i}{2} + \frac{1}{2} + B_i(r_i, Q_i)) + k \frac{\lambda_i}{Q_i} \quad (2)$$

Subject to:

$$\sum_{i=1}^N \frac{\lambda_i}{\lambda} (1 - A_i(r_i, Q_i)) \geq \beta \quad (3)$$

Where:

$i$  - index for parts,  $i = 1, \dots, N$ .

$c_i$  - unit cost for part  $i$ .

$r_i$  - reorder level for part  $i$ .

$Q_i$  - order quantity for part  $i$ .

$\mu_i$  - mean lead time demand for part  $i$ .

$B_i(r_i, Q_i)$  - average expected backorders for part  $i$ .

$A_i(r_i, Q_i)$  - long run probability of stockout for part  $i$ .

$\lambda_i$  - annual demand rate for item  $i$ .

$\beta$  - aggregate fill rate target.

The model is similar to one proposed by Hopp, Spearman, and Zhang (1997). Here the order quantity and reorder levels are functions of the aggregate fill rate target, as well as part cost, replenishment lead-time, and part demand rate. We note that multi-echelon and dynamic complexities are ignored in this model. Its purpose is to illustrate the basic tradeoff between reducing inventory costs and meeting the aggregate service target.

We wish to compare this model's performance against the actual performance revealed by our data. If the performances are similar, this would give further credence to the conjecture that DLA is focusing primarily on cost and aggregate performance, rather than the tailored performance of essentiality and criticality categories. To simplify the analysis, we aggregate our data (of 170,000 parts) into 12 groups with common essentiality and criticality codes. So instead of solving the model with 170,000 parts we aggregate them into 12 groups according to WSIC codes A through L (see Table 1) and use attribute averages within each group to characterize a 'typical' part. Although this is a very aggregate simplification of our data, it helps us solve the model efficiently while still drawing insights about the tradeoffs involved. In the real problem, each part will have its own service level and cost based on its individual characteristics. One can leverage these differences by allowing some flexibility in

how service is allocated to parts within a category, while still meeting that category’s overall service requirement. Since our aggregate model has less flexibility in differentiating service levels for parts based on their individual characteristics, we believe it is conservative (i.e., overestimates cost and underestimates service) relative to the real problem.

| System Criticality Code |        | Part Essentiality Code |      |        |     |
|-------------------------|--------|------------------------|------|--------|-----|
|                         |        | Very High              | High | Medium | Low |
| actual observed         | High   | 93%                    | 71%  | 85%    | 80% |
| descriptive model       |        | 94%                    | 74%  | 76%    | 56% |
| prescriptive model      |        | 93%                    | 90%  | 87%    | 84% |
| actual observed         | Medium | 91%                    | 77%  | 82%    | 77% |
| descriptive model       |        | 88%                    | 82%  | 76%    | 65% |
| prescriptive model      |        | 87%                    | 84%  | 81%    | 78% |
| actual observed         | Low    | 88%                    | 90%  | 80%    | 73% |
| descriptive model       |        | 82%                    | 87%  | 67%    | 61% |
| prescriptive model      |        | 84%                    | 81%  | 78%    | 75% |

Table 9: Comparison of Observed Fill Rates and Fill Rates for Descriptive and Prescriptive Models across Essentiality/Criticality Codes

Table 9 shows the predicted fill rates from our descriptive model and the actual observed fill rates across the 12 WSIC categories. Here we used an aggregate service level constraint of 91% to match the actual aggregate service observed in the data study, a  $h$  value based on a 24% annual holding cost rate, and a  $k$  of \$50. For each WSIC category, the observed fill rate is stated on the top followed by the fill rate dictated by this ‘aggregate model’ just below. For example, the observed fill rate for the most critical WSIC code is 93% while its predicted value (based on our descriptive model) is 94%. The estimated annual cost of our descriptive model is \$26.1 million, while the estimated cost of the current policy based on observed fill rates is \$27.3 million.

Table 9 offers two interesting observations concerning the impact of criticality and essentiality codes on fill rates. First, the fill rate trend across priority categories is similar

for the predicted and observed data. To see this more clearly, consider the sign of fill rate changes for adjacent criticality and essentiality categories. For example, in the first row of the table both the predicted and observed fill rates decrease as one moves from Very High to High essentiality. There are 17 such sign comparisons in the table. The predicted and observed fill rates follow the same sign change in 14 of the possible 17 comparisons. This consistency is statistically significant ( $p < 0.01$ ), using a binomial test. Our model therefore appears reasonably consistent in predicting DLA’s relative behavior across criticality and essentiality codes. It is interesting to note that in the three cases where the sign changes in observed fillrates do not match with our model, relative part cost appears to play a dominant role. For example, for low essentiality code items, the unit cost is slightly higher for high criticality items than for medium criticality items. Since our model minimizes cost, for low essentiality items, it prescribes a higher fillrate for medium criticality items than for high criticality items, contrary to the observed fillrates.

Our second observation is that the aggregate model predicts a much larger spread of fill rate values. It is interesting that a model which ignores criticality and essentiality codes altogether actually offers a wider discrepancy of performance. For example, the fill rate spread between the highest and lowest priority categories is (94%, 61%) for the model versus the observed (93%, 73%). This suggests that DLA is not setting stocking levels solely on aggregate measures, but is also sensitive not to let the fill rates of any one part category deteriorate below an acceptable bound. Our discussions with DLA support this suggestion. In the next section we offer a model to help capture this bounding behavior as well as improve DLA’s ability to directly differentiate across priority codes.

## 4.2 A Prescriptive Model

There are several ways DLA could better align its service parts performance with customer expectations. The most obvious is to offer tailored fill rates based on individual WSIC codes. One way to do this is by adding the following constraints to the model.

$$(1 - A_i(r_i, Q_i)) \geq \beta_i \quad i = 1, \dots, 12 \quad (4)$$

where  $\beta_i > \beta_j$ , if category  $i$  is more critical than category  $j$ .

For illustrative purposes, we ran this new “tailored” prescriptive model with the  $\beta_i$ ’s set to 93% for the highest criticality category ( $\beta_1$ ) and 75% for the lowest criticality category ( $\beta_{12}$ ), with a uniform slope for the criticality categories in between. These values reflect the military’s desired maximum and minimum service levels across the priority categories. Keeping the aggregate fill rate target fixed at 91%, this tailored model estimated an inventory cost of \$27.27 million which is actually slightly less than the estimated cost of the current observed performance (but about 4% larger than the cost of the descriptive model without individual fill rate constraints). This is promising since it suggests we may be able to support a strategy of setting service levels consistent with part priority with little (or no) deterioration in cost or aggregate fill rate. Table 9 provides a comparison of the fill rates from the descriptive and prescriptive models. The prescriptive model offers a tighter spread between its maximum and minimum service levels (since this was desired by the military) as well as better consistency within each priority category.

Table 9 also provides a comparison of the fill rates from our prescriptive model and the observed fill rates in practice (rows 1 and 3 of each category). An examination of this table reveals that our model increases the service provided to some of the high criticality part

categories (e.g. high essentiality and high criticality code) and reduces the service level of some low criticality code part categories (e.g. low criticality high essentiality code). Our model increases the service level of 6 categories (most of which have high criticality) while reducing the service level for 5 categories (with lower criticality). Thus by rebalancing fill rates across these categories, a more consistent service level performance can be achieved without a significant increase in cost or deterioration of the aggregate fill rate.

These results are obviously driven by the military's choice of maximum and minimum service level targets. In general, one would like a more systematic method for setting service targets (i.e., the  $\beta_i$ 's) which allowed one to examine the cost and service trade-offs of different alternatives. Note that in some environments, the allocation of parts or products to priority categories may also be an issue. This is not the case for the military since they have explicit rules for assigning priority codes to parts. A part essentiality code is usually based on an assessment of possible consequences of a part failure to the operation of the weapon system. Thus part essentiality code is usually based on engineering and other technical specifications. The essentiality and criticality codes are assigned to parts with the notion that parts with higher priority get better service.

One simple method for setting service targets is to first specify a minimum target for the lowest criticality category and then choose a uniform difference  $\Delta$  for service levels between consecutive priority categories. For example, let the service level target for the least critical category be  $\beta_{min}$ , then the service level target for the next higher criticality category would be  $\beta_{min} + \Delta$ , and for the next higher category would be  $\beta_{min} + 2\Delta$ , and so on. We can solve the constrained service level optimization problem given by (2), (3) and (4) initially by setting  $\Delta = 0$  and estimating a lower bound on the inventory cost. We then resolve the

problem by gradually increasing  $\Delta$  and estimating the new inventory cost. We repeat this process until the increased inventory cost meets the current budget. We briefly illustrate this procedure using our data.

We first classified three priority categories by using the WSIC codes given in Table 1. Category 1 denotes the highest priority level and consists of WSIC codes A, B and E. Category 3 is the lowest priority level and consists of WSIC codes H, K and L. All other WSIC codes fall in category 2. This classification is based on what DLA has proposed to implement in the future (a policy change influenced by our analysis in this study). We set a service level target of  $\beta_{min} = 80\%$  for the third category,  $\beta_{min} + \Delta$  for the second category and  $\beta_{min} + 2\Delta$  for category 1. We set  $\Delta = 0$  initially and solve the optimization problem (2), (3) and (4) resulting in a estimated cost of \$21 million. The  $\Delta$  is increased gradually until the model gave an estimated cost of \$27.3 million (the current estimated cost as discussed earlier) at  $\Delta = 5.5\%$ . Thus by accepting a service level of 80% for the category 3 parts, the service level target for the highest category parts could be increased to 91% and also have an estimated cost within the current expenditure of \$27.3 million. DLA is currently proposing to move in a similar direction by setting service level targets of 84%, 86% and 88% for the three criticality categories. They will now allocate inventory based on the part cost and the explicit service level associated with that part's priority category.

Another possible approach is to start with a service level target for the highest category and let the model optimize over the two lower service categories while maintaining a given budget target. To illustrate, we set the target for the highest service category at 95%, implying service level targets of 95%, 95% -  $\Delta$ , and 95% -  $2\Delta$  for the three categories. Setting the budget constraint to our current budget of \$27.3 million, we found an optimal  $\Delta$  of 15%.

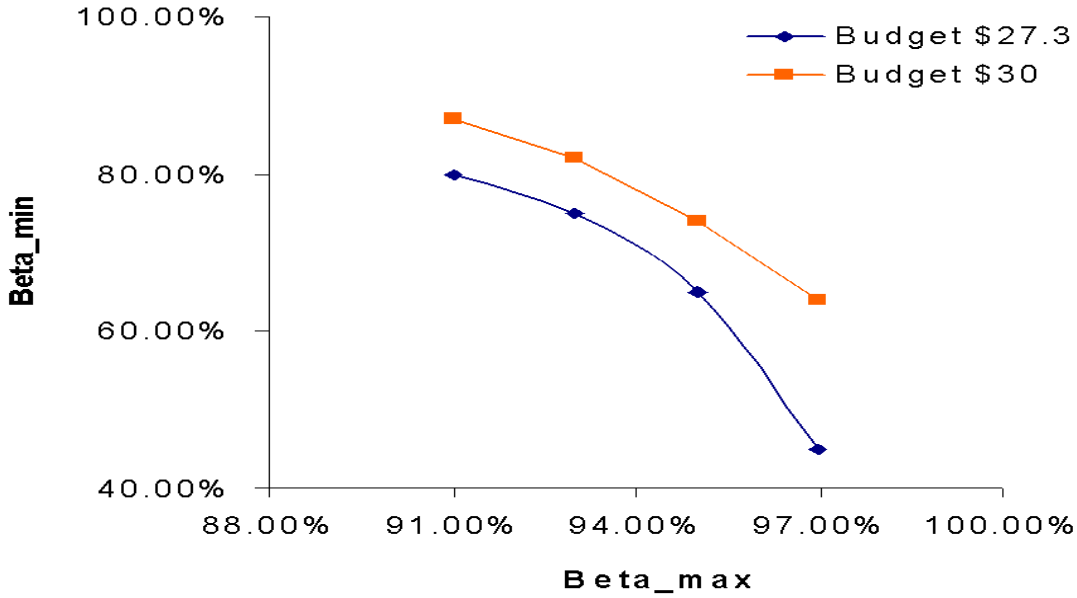


Figure 4: Tradeoff curve for  $\beta_{max}$  versus  $\beta_{min}$

Thus we would have to accept a service level of 65% on the lowest category to achieve a 95% fill rate for the highest category, while remaining within the budget. This indicates that the service levels are highly sensitive to the fill rates of the highest service category. As we increase the target for the highest service category we either have to accept reduced service levels for the lower criticality categories or have to increase the budget. This tradeoff between  $\beta_{max}$  versus  $\beta_{min}$  is shown in Figure 4. This figure shows that for very high target service levels for the highest category we may have to accept significant decrease in the service level for the low criticality categories if the budget is fixed. If the budget is increased, then the trade-off is less severe.

We would like to emphasize that these are not the only ways of setting service level targets, and indeed one could formulate this as a constrained optimization problem (see [Deshpande et al. 2001](#)). We have simply outlined two reasonable methods which appear to work well for the military. The basic idea is to set a service level target for one priority



category and then adjust the remaining targets relative to that benchmark. The critical point is that as management becomes more willing to accept a lower service level for the lowest criticality category, they can achieve higher service level targets for the highest criticality category, without significant increase in overall costs.

## 5 Conclusion

This paper reported on the role of priority codes in the military service parts system and the impact of these codes on performance. The military logistics system uses an explicit method for assigning criticality and essentiality codes to service parts. We conducted a rigorous analysis of the impact of these priority codes on performance measures such as the logistics response time in the system. Our analysis showed that due to the cost minimization objectives of the DLA under aggregate service level constraints, part cost is a significant driver of performance, whereas priority codes do not have the expected impact on performance. We constructed a descriptive inventory model to explain the trade-off between cost and service and recommend the use of explicit service level constraints based on priority categories to obtain performance consistent with the priority codes assigned to parts. We outlined two practical methods for assigning service level targets to different priority categories.

This empirical study offers a number of lessons for inventory managers in other (non-military) settings. In the commercial world, competitive success depends on customer satisfaction, of which *after sales service* is a key component. Many commercial companies provide after sales service using a complex multi-echelon network, stocking thousands of service parts. The consequences of a failure of a part to the operation of the end-product are not similar across all parts. Some parts are more critical to the operation of the end-product than others.

This paper shows the benefits of classifying parts into different priority categories based on *service*. Specifically, by classifying parts based on priority categories, the right level of service can be provided to each service category. Also, by accepting lower service levels for lower priority categories, high service levels can be provided for high priority categories without a significant increase in inventory cost. The traditional ABC classification of parts is based on part cost or part volume or a combination of both such as dollar volume. Our analysis sheds light on a customer-centric approach of classification of parts based on service requirements. Such a classification is particularly useful in managing after-sales service systems.

Our empirical findings for the DLA suggest that if the categorization of parts based on service is done based on explicit service level targets, then the differentiated service requirements can be met without a substantial increase in inventory costs. Classification of parts based on service requirement helps in achieving higher levels of service for the highly critical parts without a significant increase in overall costs. However, this classification has to be done in the following fashion. First, all parts must be properly classified according to priority based on the consequence to the operation of the end-product due to the part failure. Next, explicit service level targets should be set for each priority category.

The use of priority categories also opens up a number of interesting open research questions. For example, in this paper, we assumed a  $(Q, r)$  inventory policy was used for managing service parts inventory. The parameters for the inventory policy were set based on the fill rate target used for the priority category corresponding to the service part. The implicit assumption behind this approach is that all users of a service part assign the same priority category to that part. However, as noted in Table 6, we observed that it is quite common for different users of a specific service part to assign it different priority categories. For

example, the central agency may be supplying a common part to two customers: one with a high service requirement and the other with a low service requirement. Each customer may assign a different priority category to the same common part. The challenge for the central agency then is to be able to achieve the benefits of pooling across these heterogeneous customer classes, but at the same time provide them the differentiated service according to their requirements. Deshpande et al. (1998) analyze this rationing question in detail for a continuous review system, while Frank et al. (1999) analyze a similar problem with two priority classes for a periodic review system.

Service part categories based on service prioritization can also lead to incentive problems in a decentralized system. If parts in higher priority categories have better service then there is incentive to give a high criticality classification to as many parts as possible. During our field study we observed that the central agency was charging a fixed price to its customers independent of the service requirement required by the customer for that part. In such an environment, the ICPs could (in theory) inflate a part's priority ranking with no direct monetary consequences. The military is now exploring ways to encourage individuals in their decentralized system to assign the correct priority category, and associated service level, to each part. Deshpande et al. (2000) provide an initial analysis of this incentive problem.

## **Acknowledgments**

This research was supported in part by the U.S. Navy under Contract # NOV391-96-M-M04, NSF CAREER Award #9602072, NSF grant #0075391, and the Fishman-Davidson Center for Service Operations Management at the Wharton School. Special thanks to the following people from various organizations within the military - Sandy Leggieri, Gary Burchill, Jere

Engelman, and Mike Puoy. The authors also acknowledge the detailed comments of Alan Washburn and two anonymous referees on earlier versions of this paper.

## References

Cohen, M. A., K. Donohue, and V. Deshpande. 1998. Supply Chain Coordination Study: U.S. Navy / Defense Logistics Agency. Project Report, *Fishman-Davidson Center for Service and Operations Management*, The Wharton School, Philadelphia.

Cohen, M. A., Kamesan, P., Kleindorfer, P. R., and Lee, H. L. 1990. OPTIMIZER: A Multi-Echelon Inventory System for Service Logistics Management. *Interfaces*, 20(1), 65-82.

Cohen, M. A., P. R. Kleindorfer, and H. L. Lee. 1986. [Optimal Stocking Policies for Low Usage Items in Multi-Echelon Inventory Systems. \*Naval Research Logistics Quarterly\*, 33, 17-38.](#)

Cohen, M. A., P. R. Kleindorfer, and H. L. Lee. 1988. [Service Constrained \(s, S\) Inventory Systems with Priority Demand Classes and Lost Sales. \*Management Science\*, 34\(4\), 482-499.](#)

Cohen, M. A., P. R. Kleindorfer, and H. L. Lee. 1989. [Near Optimal Service Constrained Stocking Policies for Spare Parts. \*Operations Research\*, 37, 104-117.](#)

Cohen, M. A., and Lee, H. L. 1988. [Strategic Analysis of Integrated Production-Distribution Systems: Models and Methods. \*Operations Research\*, 36, 216-228.](#)

Cohen, M. A., and Lee, H. L. 1990. Out of Touch with Customer Needs? Spare Parts and After Sales Service. *Sloan Management Review*, Winter, 55-66.

Cohen, M. A., and Zhang, S. 1997. Benchmarking Service Parts Logistics: An In-Depth Analysis. *Working Paper*, The Wharton School, University of Pennsylvania.

Deshpande, V., Cohen, M. A., Donohue, K. 2001. A Threshold Inventory Rationing Policy for Service Differentiated Demand Classes. *Working Paper*, The Wharton School, University of Pennsylvania, August.

Deshpande, V., Cohen, M. A., Donohue, K. 2000. Incentive Compatible Pricing Mechanisms for Service Differentiated Supply Chains. *Working Paper*, The Wharton School, University of Pennsylvania, August.

Deshpande, V. 2000. Supply Chain Coordination with Service Differentiated Customer Classes. *unpublished Ph.D. dissertation*, The Wharton School, University of Pennsylvania, August.

Deuermeyer, B., and Schwarz, L. B. 1981. A Model for the Analysis of System Service Level in Warehouse/Retailer Distribution Systems: The Identical Retailer Case.in Schwarz, L. B.(ed.), *Multi-Level Production/Inventory Control Systems, TIMS Studies in Management Science*, 16, North-Holland, Amsterdam, 163-193.

Frank, K., Zhang, R. Q., and Duenyas, I., 1999. Optimal Inventory Policies in Systems with Priority Demand Classes, under revision with *Operations Research*.

Graves, S.C. 1985. [A Multi-Echelon Inventory Model for a Repairable Item with One-for-One Replenishment.](#) *Management Science*, 31, 1247-1256.

Hitt, L. M., and Frei, F. X. 1999. [Do Better Customers Utilize Electronic Distribution Channels? The Case of PC Banking.](#) *Working Paper*, The Wharton School, University of Pennsylvania, April

Hopp, W. J., Spearman, M. L., and Zhang, R. Q. 1997. Easily Implementable Inventory Control Policies, *Operations Research*, 45, 327-340.

Lee, H.L. 1987. [A Multi-Echelon Inventory Model for Repairable Items with Emergency Lateral Transshipments.](#) *Management Science*, 33, 1306-1316.

Lee, H., and Billington, C. 1993. [Material Management in Decentralized Supply Chains.](#) *Operations Research*, 41(5), 835-847.

Muckstadt, J. A., and Thomas, L. J. 1980. Are Multi-Echelon Inventory Methods Worth Implementing in Systems with Low Demand Rates?. *Management Science*, 26, 483-494.

Rustenburt, W. D., vanHoutum, G. J., and Zijm, W. H. Oct 2000. Spare parts management

for technical systems: Resupply of spare parts under limited budgets. *IIE Transactions*, (32), 1013-1026.

Sherbrooke, C. C. 1968. METRIC: A Multi-Echelon Technique for Recoverable Item Control. *Operations Research*, (16), 122-141.

Simon, R.M. 1971. Stationary Properties of a Two-Echelon Inventory Model for Low Demand Items. *Operations Research*, 19, 761-777.

Wang, Y., Cohen, M.A. and Zheng, Y, S. 1999. Identifying Opportunities for Improving Teradyne's Service Parts Logistics System. *Interfaces*, (29), 1-18.

Wang, Y., Cohen, M.A. and Zheng, Y, S. 2000. A Two-Echelon Repairable Inventory System with Local-Center Dependent Depot Replenishment Lead Times. *Management Science*, (49), 1441-1453.

Wang, Y., Cohen, M.A. and Zheng, Y, S. 2001. Differentiating Parts Replacement Service on the Basis of Delivery Lead-Times. Forthcoming, *IIE Transactions*.