Multi-Echelon Multi-Product Inventory Strategy in a Steel Company

By: Juan D. Iocco
Thesis Advisor: Amanda J. Schmitt, PhD

Summary: Ternium is an integrated steel mill that faces customers’ requirements of short service times and high service levels. However, increasing safety stock is not a desired solution. This project addresses the problem of determining where to allocate inventories in a long lead-time multi-echelon system, and its impact on inventory costs and service level.

Juan Iocco will continue work in Supply Chain in Ternium, a South American steel making company. His last positions in this company were: Supply Chain Regional Director, Production Planning and Scheduling Manager and Order Management Manager.

KEY INSIGHTS
1. The right inventory allocation drives savings in holding costs in multi-echelon production systems.
2. Simple practices, such as applying single-echelon inventory policies - and in particular holding safety stocks just in the first echelon - produce high-cost results in this case.
3. Simulation is a good tool to test and compare different policies, practices and find efficient frontiers.

Introduction
The steel production process involves up-to 10 facilities and a lead-time that ranges from 3 weeks to 8 weeks for different products. This project focuses on items with long lead times and short service times (less than a week). Common sense suggests carrying more inventories and locating them close to the customers. However, steel company’s strategies rely on low costs; holding costs associated with carrying high volumes of safety stocks are expensive. In addition to financial costs, there are high costs associated with the lost of value of aged items.

Based on the service time required and the industry’s practices, this study is focused on the last three production system’s levels (or echelons). The products at these echelons are: cold-rolled steel coils, coated cold-rolled steel coils and customized coated steel. The system and the approximated number of SKU involved at each level are depicted in Figure 1.

Demand is satisfied from the first echelon. The number of SKU items change because the items are aggregated at each upstream echelon, setting-up what is called a family of products. Each one of these families can be represented as a different 3-echelon distribution network. The most simple of them is a serial system (only 3 facilities), the most
complex ones are distribution systems with several facilities at every echelon. A model of 1, 3 and 5 items, at echelons 3, 2 and 1 respectively, was selected to analyze the performance of the system (Figure 2).

![Figure 1: Three-Echelon System.](image)

Finally, the questions that this project answers are where inventory should be held in Ternium's supply chain, and how customer service level, inventory holding costs and percentage of bounded demand affect the solution. In addition, we propose to analyze if simple methods can lead to good economical results in this multi-echelon problem.

**Method**

Distribution multi-echelon systems do not have an analytical closed-form optimal solution. That makes the choice of the method to calculate the safety stocks non-trivial. Based on this fact and on the need to perform sensitivity analysis and test different methods, we selected simulation to perform the main analysis.

A simulation model was constructed in Excel. It runs a base-stock model for the distribution system of Figure 2 for a selected typical item family. The model runs by simulating the system for more than 1,300 inventory allocations and finding the best (lowest-cost) results.

The company does not have a clear model for holding costs. Therefore three different holding costs scenarios were evaluated based on company’s data.

In addition, we tested three alternative methods that can be used to determine the base-stock levels, too. The first of them considers that the inventory information is not shared in the system and decentralized decisions are made based on local conditions (a single-echelon inventory model); the other two methods are centralized and assume that inventory information and demand is shared in the system (an echelon-inventory policy and Graves-Willem method). All the models described (including simulation) apply base-stock policies. That means that a base-stock level is set for every facility and every period of time, after satisfying demand, each facility reorders material to sep-up again the inventory position to the base-stock level.

In the single-inventory model, the decision-maker selects where to allocate the inventories and the model is used to determine the base-stock levels. In this case 3 strategies are run: inventory close to the customer (S-E 1), inventory far from the customer (S-E 2) and an intermediate one (S-E 3).

In the centralized decision models, both the inventory volume (base stock-levels) and locations are decided. The echelon inventory policy (EIP) uses the concept of echelon inventory (calculating the base-stock level at every level considering the system as all the downstream echelons and inventory from the customer to the analyzed level). Figure 3 shows the 3 strategies for the single-echelon model and the applied the concept of echelon inventory position.

The last model, the Graves-Willems method, used a dynamic programming algorithm to calculate the base-stock level as a function of the service time that every facility provides to the next one. It was applied using the software, “Power Chain Inventory”.

Finally, all the base-stock levels calculated with these alternative methods were tested in the simulation model in order to compare the outputs with the best results.
Figure 3: Single-Echelon and Echelon inventory Model. Each echelon contains more than one SKU.

Results

First, the three cost scenarios were simulated and the lowest cost solutions which satisfy a minimum 95% service level were selected for each scenario. Table 1 shows the results. The $k_{sim}$ parameters represent the amount of safety-stock inventory allocated at each echelon, in terms of a function of demand standard deviation.

Table 1: Simulated Best Solutions for 95% Service Level.

<table>
<thead>
<tr>
<th>Cost Scenario</th>
<th>1</th>
<th>2</th>
<th>3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Echelon</td>
<td>3</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>Cost pattern</td>
<td>Low</td>
<td>Low</td>
<td>Low</td>
</tr>
<tr>
<td>$k_{sim}$</td>
<td>0.5</td>
<td>2.4</td>
<td>1.7</td>
</tr>
</tbody>
</table>

Table 1 shows that safety stock allocation changes depending on the cost scenario. It is allocated close to the customers (higher value $k_{sim}$ parameter at facility 1) when all then facility costs are approximately equal, and up-stream in the system when holding cost is more expensive in the first facility (increase in value of $k_{sim}$ parameter at facilities 2 and 3).

Next the complete set of simulated results was used to find the lowest holding cost for each service level. These values set up what it is called the efficient frontier. In addition, the calculated base-stock policies of the alternative methods were simulated and graphed with these results. The efficient frontier for the first cost scenario is shown in Figure 4. The other two are not shown because of space constraints, but the alternative models perform similarly. All the holding cost results shown in this work correspond to a selected family of SKU selected as example.

We also tested scenarios of “bounded demand”. These scenarios represent agreements with customers to limit their actual orders to be only a percentage above their forecasted mean (instead of any level of variation). Figure 5 shows that holding costs are reduced meaningfully in this case; the green line shows holding costs when demand is unbounded, the blue shows it bounded to 50% above the mean, and the red to 100% above the mean.

Another topic of study was on-hand inventory. It is a common practice to measure inventory performance in terms of total on-hand inventory. Figure 6 depicts the total on-hand inventory for the efficient frontier solutions given in Figure 3. We see that for the best holding cost solutions the curve is not continuous. This means that is possible, in some cases, to increase service level while decreasing on-hand inventory.
Finally, Table 2 shows how different methods perform compared with the best solution at 95% service level. It gives the average difference, in percentage of holding costs, of each method against the best solution (the average considers the difference for all three cost scenarios).

Table 2: Simulation of Alternative Methods. Cost Variation of alternative methods vs. the best results.

<table>
<thead>
<tr>
<th>Alternative Methods</th>
<th>% Variation of Holding Cost</th>
</tr>
</thead>
<tbody>
<tr>
<td>Echelon Inventory Model</td>
<td>8.2%</td>
</tr>
<tr>
<td>Graves-Willems Model</td>
<td>9.7%</td>
</tr>
<tr>
<td>Single-Echelon 1</td>
<td>35.0%</td>
</tr>
<tr>
<td>Single-Echelon 2</td>
<td>15.4%</td>
</tr>
<tr>
<td>Single-Echelon 3</td>
<td>18.5%</td>
</tr>
</tbody>
</table>

Conclusions

The mains conclusions of this work are:

- The calculation of the safety stock allocation can be accomplished using simulation, but results are very sensitive to cost parameters.

- Focusing solely on the control of on-hand inventory levels, or cost of on-hand inventories, leads to inefficient results.

- Single echelon models do not perform as well as stand-alone multi-echelon methods.

- Keeping all the safety stock just in the first echelon, close to the customer, is a bad solution with respect to cost. In this research case, holding cost increased by 35%.

- The Echelon Inventory Policy and Graves-Willems models performed well.

- Customer contract variables could be important in the solution; bounding the demand can significantly lower required inventory levels, and hence holding costs.

- An in-depth characterization of the product and demand must be the first focus of these inventory problems. Based on these characterizations, policies can be set up with the methods suggested in this work.

- IT planning systems must be prepared to support different allocation strategies, including different possibilities between Make-to-Order and Make-to-Stock.