Inventory Management

Duvel Case Study

January 31, 2025

Contents

Introduction

Our mission is straightforward: help Duvel figure out which forecast works best and analyze how it impacts different inventory management strategies.

To do this, we'll test various stock policies and assess the forecast using key performance indicators that truly matter.

Let's dive in!

Data Preparation

When it comes to analysis, data quality is everything—it directly impacts the reliability of our results. As the client pointed out, there are missing values in both forecasts for weeks 36, 37, and 38 of 2023.

There are several ways to handle missing forecasted values, and we considered the following approaches:

- **Interpolation:** This works best when data points change smoothly over time, which isn't quite the case here.
- **Regression models:** A solid option, but a bit too complex given that we don't have details on how the statistical forecast was generated. Using a regression could increase accuracy unnecessarily, making it more sophisticated than the other methods we're comparing.
- **Using actual sales data:** While this would act as a neutral equalizer, it would slightly inflate key performance indicators, making the forecasts appear more accurate than they actually are.
- **Moving average (last three weeks): This approach keeps things simple and fair—each forecast is filled with the same values, maintains some variability with real sales, and relies on a well-known, easy-to-apply forecasting method. Plus, moving averages are surprisingly strong predictors.**

Given that the missing values account for only about 4% of the dataset, and our main goal is to compare forecasts rather than perfect them, the moving average method strikes the right balance. It avoids adding unnecessary complexity, ensures neither forecast is favored, and keeps accuracy largely unchanged.

Weeks 45/46/47 2024

For my specific product, I also noticed that the statistical forecast for weeks 45, 46, and 47 of 2024 was set to zero.

At this point, I had to decide whether to apply the same criteria as before or leave it as is. I went with the latter, and here's why:

- The client didn't explicitly mention these values.
- They're not actually zero but rather an extremely small number (10^{10}), likely the result of some calculation.

This suggests a poor forecast from the statistical tool, and instead of adjusting for it, I felt it was important to recognize and account for this inaccuracy.

In short, leaving these values untouched ensures that the forecasting method is fairly assessed; with flaws included.

Key Performance Indicators

It's important to define the KPIs we'll use to optimize and evaluate the performance of the models.

For our study, we'll focus on three main KPIs. The Bias, Bias $= \frac{1}{n} \cdot \sum e_t$, which measures the average difference between predicted and actual values, indicating systematic error. The Mean Absolute Error, MAE $= \frac{1}{n} \cdot \sum |e_t|$, which captures the average magnitude of errors, providing a straightforward measure of accuracy. And the Root Mean Squared Error, RMSE $= \sqrt{\frac{1}{n} \cdot \sum e_t^2}$, which emphasizes larger errors due to its sensitivity to outliers.

$$\text{Bias }\% = \frac{\sum e_t}{\sum sales}$$

$$\text{MAE }\% = \frac{\sum |e_t|}{totalsales} \cdot \sum |e_t|$$

$$\text{RMSE }\% = \sqrt{\frac{1}{n} \cdot \sum e_t^2}$$

We'll combine these KPIs to take a more detailed approach and achieve better optimization of our models. Specifically, we'll optimize the models using two metrics, the score and the sum:

$$\text{Score} = |\text{Bias}| + \text{MAE}$$

$$\text{Sum} = |\text{Bias}| + \text{MAE} + \text{RMSE}$$

This approach is interesting because it strikes a balance between bias and accuracy across different models while also accounting for sensitivity to outliers through the RMSE. Since no single metric is perfect, this method lets us leverage the strengths of several combined.

Forecast Comparison

Using these KPIs, we obtained the following results:

	MAE	RMSE	Bias
Statistical Forecast	44.79%	68.35%	-6.25%
Consensus Forecast	45.15%	66.34%	2.30%

From this, we can calculate the Score and Sum metrics:

	Score	Sum
Statistical Forecast	51.04%	119.39%
Consensus Forecast	47.45%	113.80%

A key takeaway here is that one of the biggest weaknesses of the statistical forecast is its bias, which exceeds 6%. While we don't have details on how this forecast is generated, this level of bias suggests that a closer examination is needed.

Overall, the consensus forecast performs better across almost all metrics. With a lower score and sum, it appears to be the more accurate option.

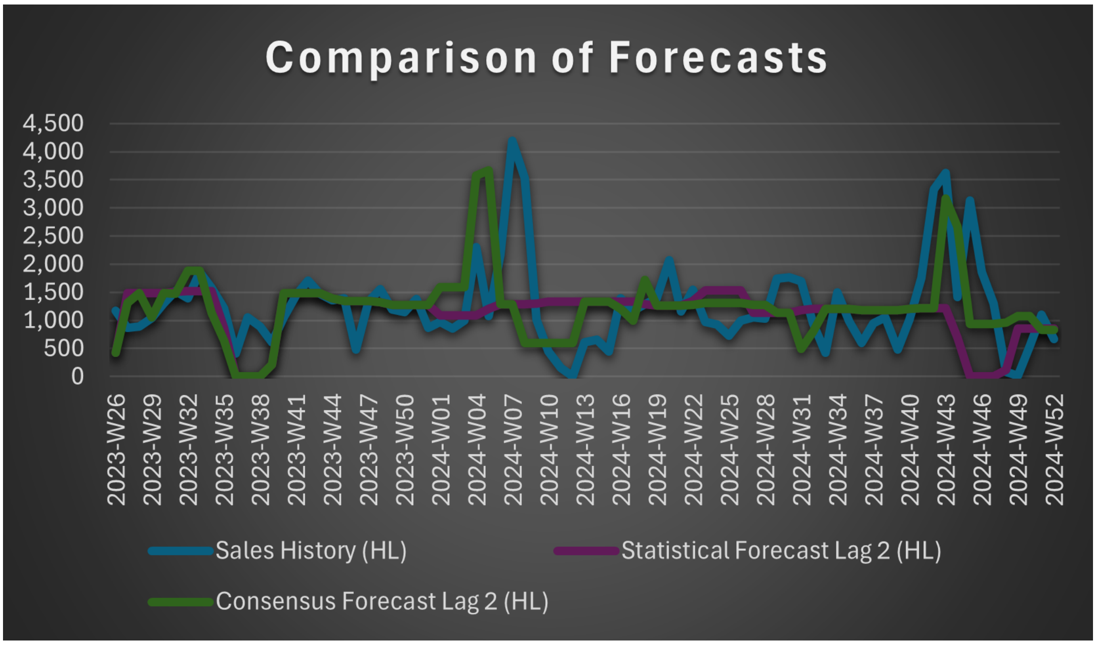

Figure 1: Forecast Comparison

Additionally, the graph shows that while the consensus forecast isn't perfectly timed, it does capture some demand spikes—something the statistical forecast largely misses, as it remains quite flat. This is a promising sign and could be due to additional information being fed into the model that the statistical forecast doesn't account for.

Comment on the Sales History

When creating a forecast, we should always focus on demand, not sales. Otherwise, we risk falling into a loop: if we forecast zero sales, we won't produce anything, which in turn leads to selling nothing, reinforcing the cycle.

For this case study, only the sales history was provided. However, to proceed with the analysis, I will assume that these sales represent the actual orders placed for this specific product by Duvel—essentially serving as a proxy for demand.

Stock Policies

For this study, we will implement and test four different stock policies and compare their effects on the two forecasts we have.

Before diving into the specifics of each policy, it's important to establish the criteria we'll use to evaluate them.

Given the lack of information regarding MOQs, bulk order discounts, or any other purchasing constraints, we will assume that we can order any quantity we want, as long as we adhere to the lead time and review period.

Moreover, ideally, any comparison between models should be expressed in financial terms. If improving forecast accuracy by 5% costs €10 but doesn't change purchasing decisions due to supplier MOQs, then it's not an improvement—it's just a €10 extra cost.

However, in this case, we don't have financial data, making it difficult to conduct a precise cost-based analysis. While we could attempt to estimate costs using online sources, the margin of error would likely be so large that it would undermine the entire study.

This is why we'll focus on two key metrics: fill rate and average inventory.

- **Fill rate:** The percentage of lost sales relative to total sales over the study period.
- **Average inventory:** The average stock level maintained in the warehouse.

In general, increasing the fill rate requires holding more inventory, so the challenge is finding the right balance.

The beer market is highly competitive, with numerous players offering similar products. Given this, it's reasonable to assume that stock shortages result in lost sales—customers will simply switch to competitors if a product isn't available. Additionally, stockouts mean missed opportunities to strengthen brand reputation and increase market share. This makes maintaining a high fill rate a major advantage.

That said, we can't overlook the downsides of holding excessive inventory, which include:

- **Higher storage costs:** Warehousing isn't free, and maintaining large stock levels increases expenses for space, utilities, and labor.
- **Risk of obsolescence:** Beer has a shelf life, and holding too much stock increases the chances of products expiring or becoming unsellable.
- **Tied-up capital:** Excess inventory means money that could be invested elsewhere is instead locked in stock. This can hurt cash flow and limit financial flexibility.

Among many others.

As a final note, to ensure a fair comparison between policies, I've set the starting inventory at 5,000 units for every scenario. While this might seem high, it's necessary given our two-week lead time—meaning no new stock will arrive in the first week. This initial buffer helps account for that gap and ensures a smoother evaluation of each policy.

Now, let's move on to analyzing each policy!

Policy 1: Static Replenishment

To kick things off, we assessed a simple, static replenishment policy. Under this approach, we review inventory levels weekly and order just enough to bring stock back up to 4,000 units, with a 2 week lead time.

Since this policy operates independently of any forecast, the results are identical for both forecasting methods.

Results Summary

Metric	Value
Fill Rate	90.91%
Monthly Service Level	88.61%
Average Inventory	2,510 units

Table 1: Performance Metrics - Policy 1

This policy serves as a baseline for future comparisons. Without a reference point, it's hard to judge whether a given fill rate is sufficient, how much inventory we should realistically hold, and where the trade-offs lie.

That said, as we've discussed before, the competitive nature of the beer industry pushes us toward high fill rates—ideally above 90

Observations

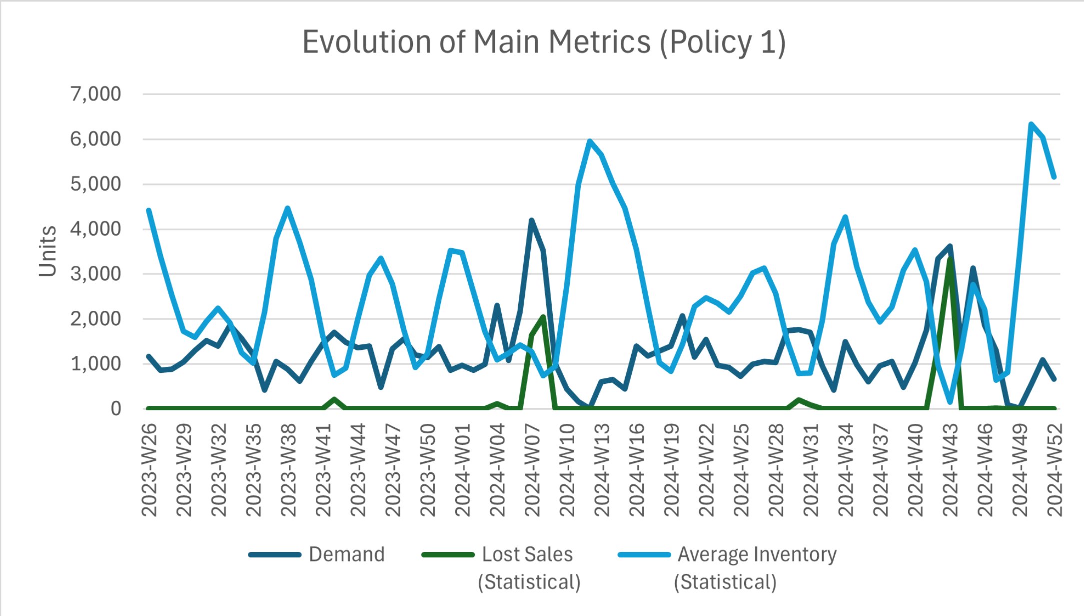

Figure 2: Evolution of Key Metrics - Policy 1

From the graph, we can clearly see a spike in lost sales, immediately followed by an increase in average inventory. This happens due to our two-week lead time—orders placed to restock the missing inventory take time to arrive, causing delays in meeting demand.

While this policy is simple and easy to implement, it doesn't leverage any predictive capability. Given that we're operating in a demand-driven industry, we expect to see significant improvements with more sophisticated stock policies.

Let's move on and see how we can do better!

Policy 2 - Replenishment with Dynamic Up-to Level

For this policy, we maintained a replenishment approach but introduced a dynamic element. Instead of a fixed target, we now set our up-to level based on the forecast. Specifically, we take the sum of the next three months' forecasted demand and add a fixed safety stock of 400 units. In other words, at each review period, we check how much we need to reach this threshold and order the difference.

Since our replenishment now depends on the forecast, we expect to see differences in performance between the two forecasting methods. Let's dive into the numbers:

Metric	Statistical Forecast	Consensus Forecast
Fill Rate	87.84%	93.49%
Monthly Service Level	88.61%	93.67%
Average Inventory	2,542	2,604

Table 2: Performance Comparison - Policy 2

We can clearly see the superiority of the consensus forecast, achieving a remarkable 93.49% fill rate while keeping almost the same average inventory as the statistical forecast.

Comparison with Policy 1

Compared to our baseline (Policy 1), this approach brings notable improvements. The consensus forecast, for instance, manages to increase the fill rate by 2.6 percentage points while only slightly raising the average inventory (by about 3.7%). That's an efficient trade-off!

Now, let's take a look at how this policy behaves over time:

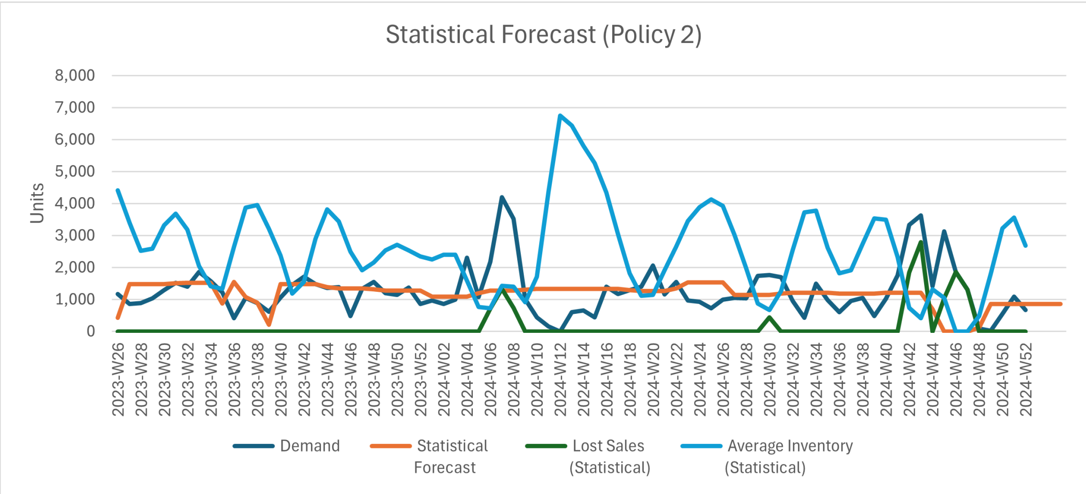

Figure 3: Evolution of Key Metrics - Statistical Forecast (Policy 2)

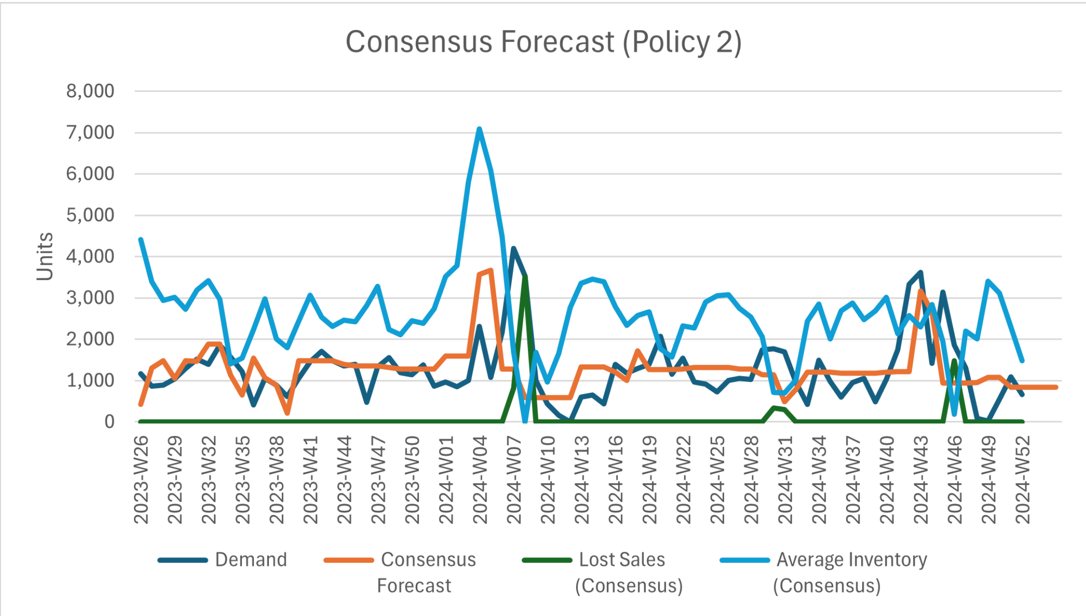

Figure 4: Evolution of Key Metrics - Consensus Forecast (Policy 2)

As expected, both policies struggle with demand spikes. However, the consensus forecast allows us to anticipate them more effectively, reducing lost sales.

A Key Drawback

A major downside of this policy (which, to be fair, is present in all of them) is that inventory levels can skyrocket at certain moments.

For instance, we see a peak of nearly 7,000 units at one point. A closer look at the data reveals that at the beginning of that month, inventory exceeded 8,000 units. This kind of excessive buildup can create serious operational challenges:

- Storage capacity may become overwhelmed.
- Warehouse staff might struggle to handle the workload, leading to outsourcing.
- Outsourcing drives up costs immediately.

Without more financial data, it's difficult to fully quantify the impact, but it's an important factor to keep in mind when designing stock policies.

Policy 3 - Fully Dynamic Replenishment

Setting a fixed 400-unit safety stock was a bit arbitrary, so we're taking a different approach. Instead of a static buffer, we'll define our up-to level as three months of forecasted demand plus percentage of the fourth month's forecast (our new dynamic safety stock). This percentage, or coverage factor, determines how much extra stock we hold as a buffer against uncertainty.

To evaluate the impact of this change, we simulated eight different coverage levels (ranging from 0 to 0.7) and analyzed how they affect our fill rate and average inventory.

Results for a Coverage Factor of 0.3

Let's start by examining the performance of a 0.3 coverage factor:

Metric	**Statistical Forecast**	**Consensus Forecast**
Fill Rate	87.05%	93.57%
Monthly Service Level	84.81%	91.14%
Average Inventory	2,525	2,851

Table 3: Performance Comparison - Policy 3 (Coverage 0.3)

Again, the consensus forecast outperforms the statistical forecast, achieving a higher fill rate. However, we now require slightly more inventory to reach a similar fill rate as Policy 2. This suggests that while dynamic safety stock provides flexibility, it may not always lead to greater efficiency.

Impact of Coverage on Fill Rate and Inventory

By plotting fill rate vs. average inventory for different coverage levels, we obtain the following relationship:

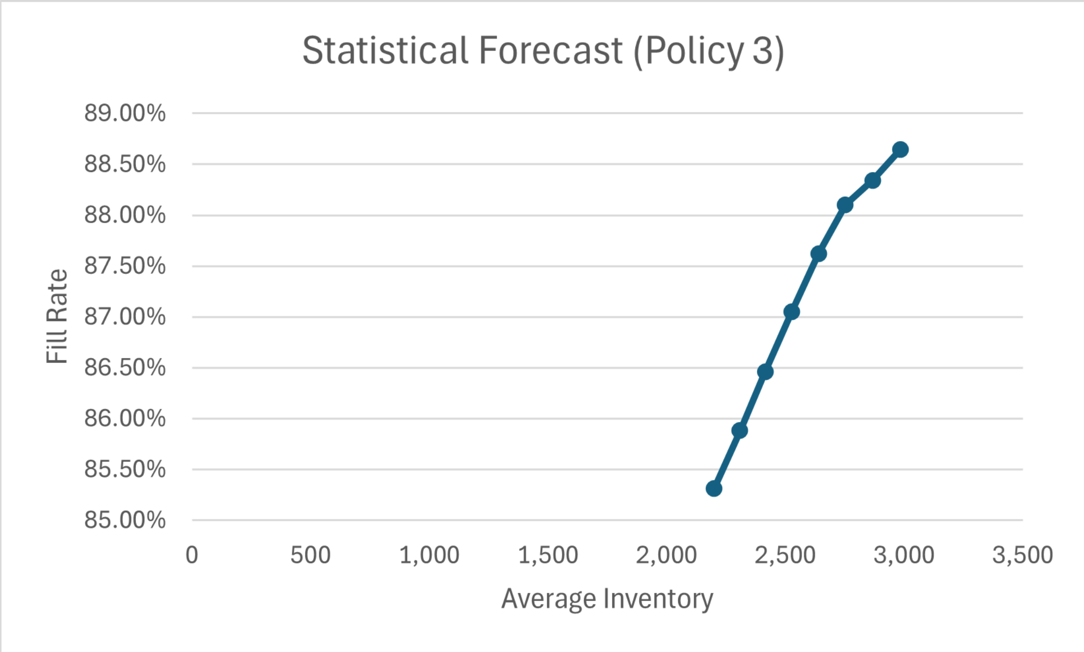

Figure 5: Fill Rate vs Average Inventory - Statistical Forecast (Policy 3)

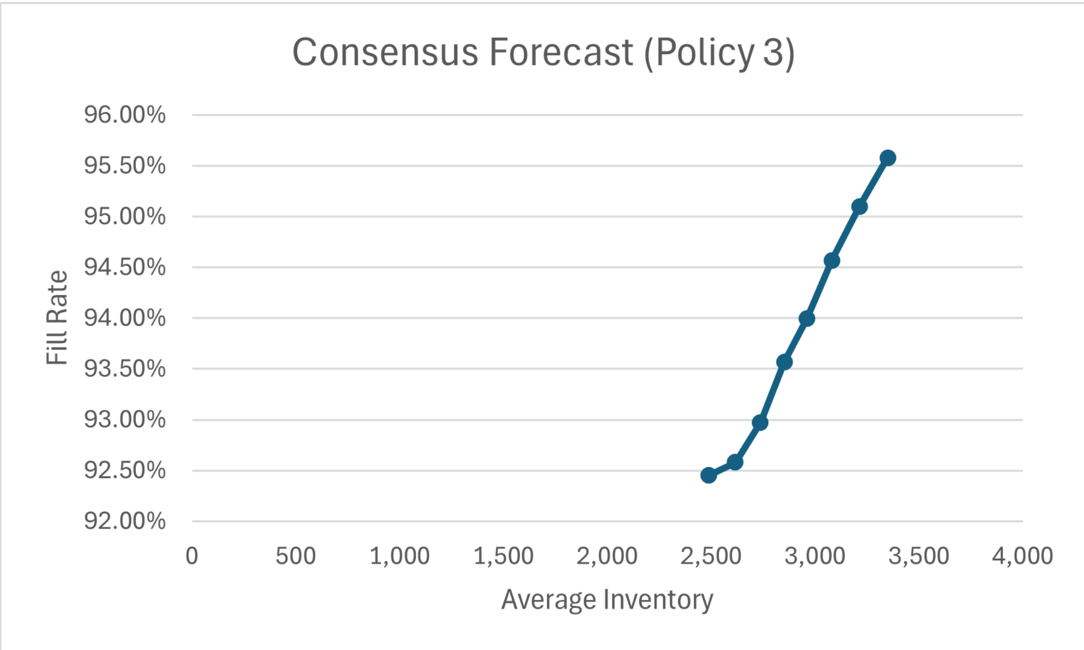

Figure 6: Fill Rate vs Average Inventory - Consensus Forecast (Policy 3)

The results show a clear positive and linear relationship between fill rate and inventory. In other words, higher inventory leads to better service levels, but the efficiency of this trade-off diminishes as coverage increases. Moreover, when increasing coverage above 0.7 there's a tipping point where increasing coverage barely improves fill rate but drastically inflates inventory, leading to higher holding costs.

Key Takeaways from the Evolution of Metrics

Now, let's examine how this policy performs over time:

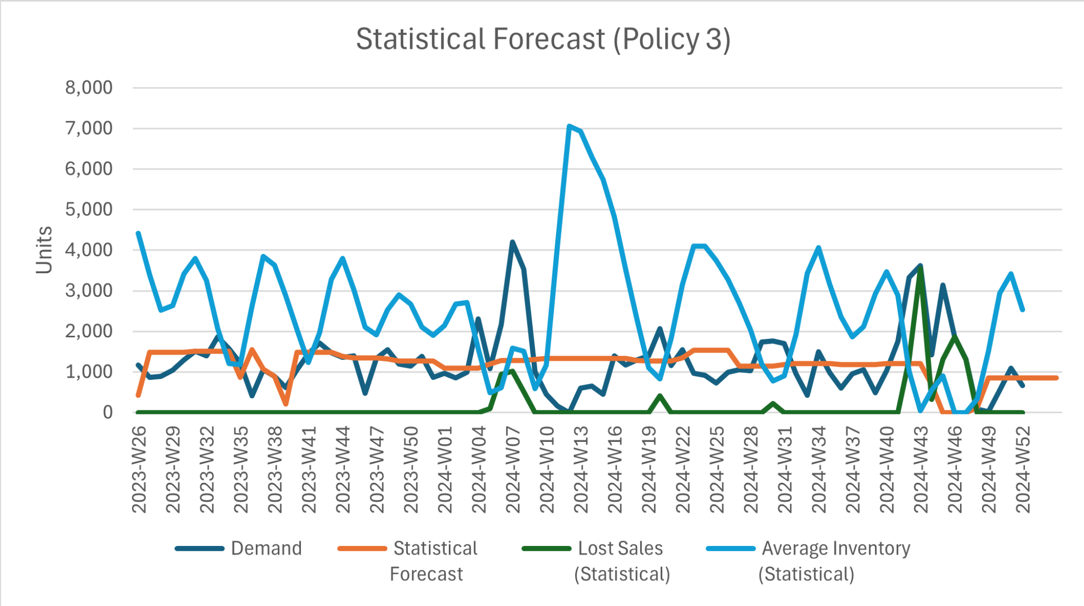

Figure 7: Evolution of Key Metrics - Statistical Forecast (Policy 3)

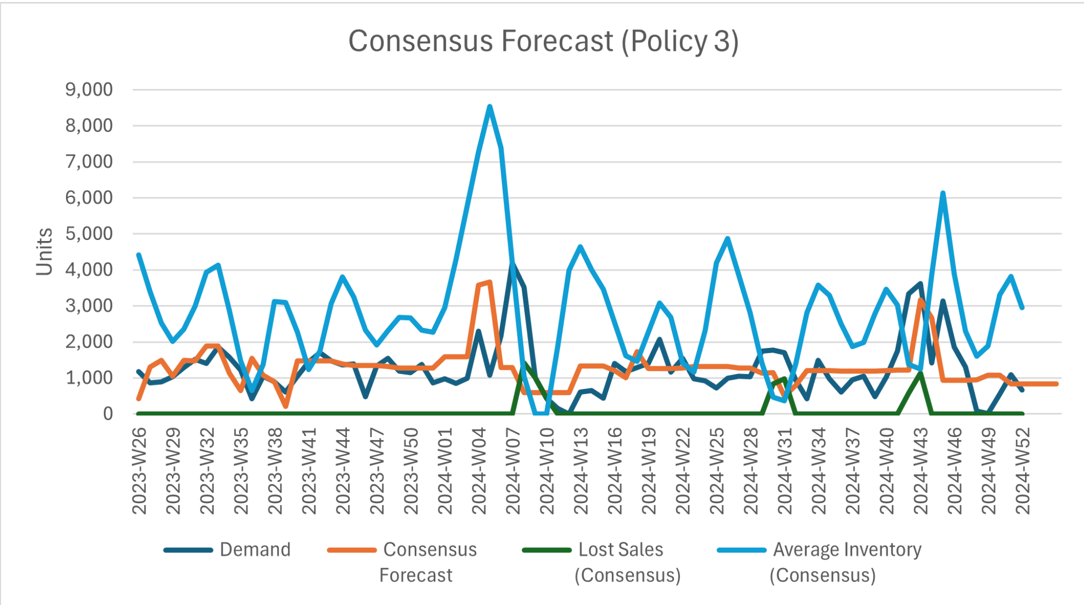

Figure 8: Evolution of Key Metrics - Consensus Forecast (Policy 3)

Here, we observe two major issues:

1. Excessive Inventory Spikes. As seen in previous policies, inventory surges to extreme levels at certain points.This can lead to operational bottlenecks, storage constraints, and higher costs due to possible outsourcing.
2. Zero Inventory Situations At some points, average inventory drops to zero, resulting in completely lost sales for that week. This highlights a critical flaw: forecasts are not always reliable, and basing replenishment exclusively on them can backfire.

Final Thoughts on Policy 3

While making safety stock dynamic provides more flexibility, it comes with trade-offs. On the one hand, it's more adaptable to demand changes and has a better balance between service level and inventory (compared to fixed safety stock). On the other hand, still struggles with forecast inaccuracies and can lead to extreme fluctuations in inventory levels.

To mitigate the risks, we might need to introduce constraints or apply a hybrid approach, where safety stock is partially based on historical variability rather than just forecast projections.

Policy 4 - Historical Variability-Based Safety Stock

For our final policy, let's take a more adaptive approach. Instead of setting a fixed safety stock, we'll dynamically adjust it based on historical forecast errors. Each month, we compute the Root Mean Square Error (RMSE) over the past 12 months and use this to determine the safety stock, factoring in the desired service level, the lead time and the review period.

The formula we apply is:

$$SS = z \cdot RMSE \cdot \sqrt{T}, \tag{1}$$

where T is the risk horizon (lead time + review period), which in our case is 3 weeks.

To evaluate this policy, we tested nine different service levels: 50%, 60%, ..., 90%, 95%, 96%, 98%, and 99%. Below, we analyze the results for a 90% weekly service level:

Metric	Statistical Forecast	Consensus Forecast
Fill Rate	92.79%	95.83%
Monthly Service Level	88.61%	94.94%
Average Inventory	3,645	3,936

Table 4: Performance Comparison - Policy 4 (90% Service Level)

Analysis of Results

This policy leads to a significant increase in inventory, which translates into a higher fill rate. No surprises here—higher safety stock provides better coverage against demand variability. However, we can already spot a major drawback: compared to Policy 3, we need more inventory to achieve similar fill rates. This suggests that relying purely on historical forecast errors may not be the most efficient approach.

As usual, the consensus forecast outperforms the statistical forecast, achieving a higher fill rate with a slightly higher inventory cost.

To better understand how this policy compares with previous ones, let's examine the relationship between fill rate and inventory levels for different service levels.

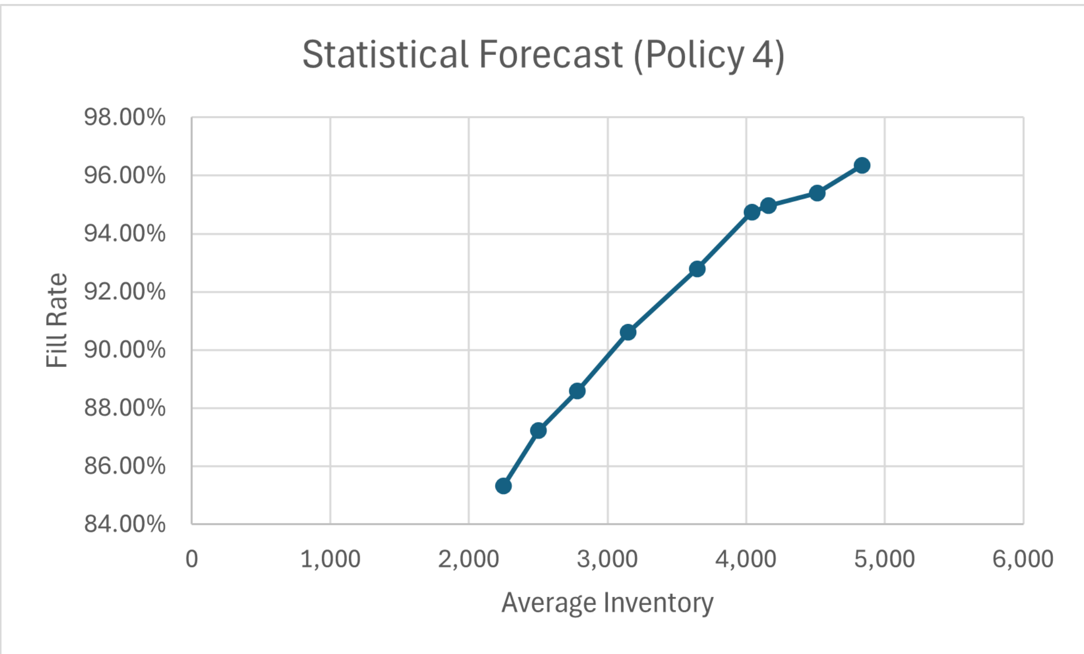

Figure 9: Fill Rate vs Average Inventory - Statistical Forecast (Policy 4)

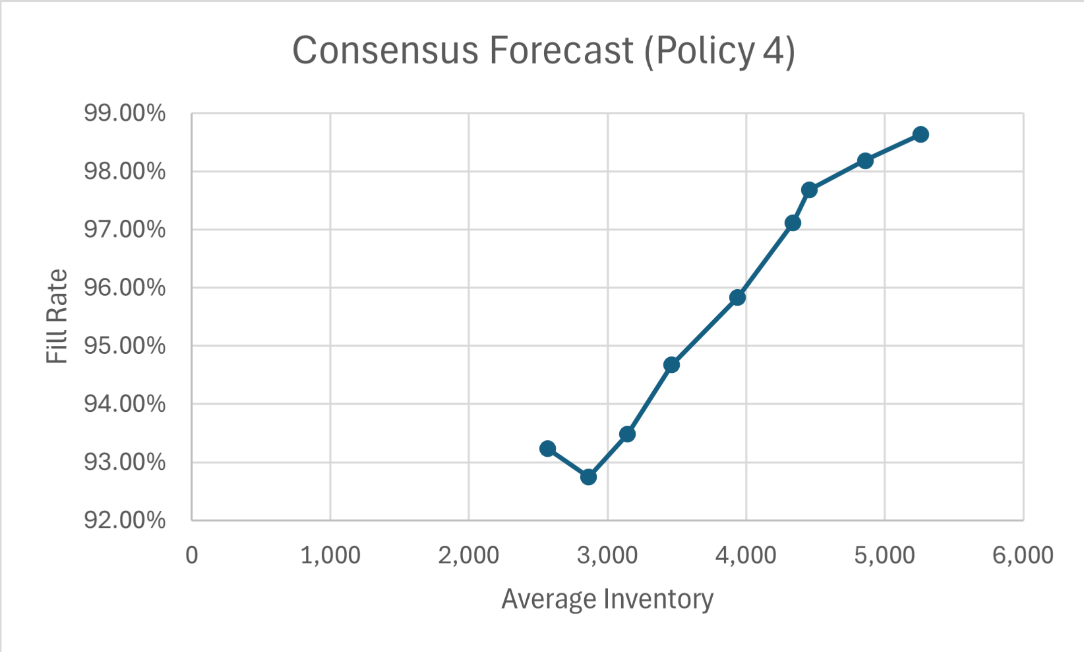

Figure 10: Fill Rate vs Average Inventory - Consensus Forecast (Policy 4)

What's Different?

The relationship between fill rate and inventory remains linear, reinforcing the fundamental trade-off: to improve service levels, you need more stock. However, something stands out: Policy 3 achieved the same fill rate with less inventory. This suggests that a fixed safety stock (based on future demand) was actually more efficient than one based on historical errors.

Let's take a deeper look at how key metrics evolved over time.

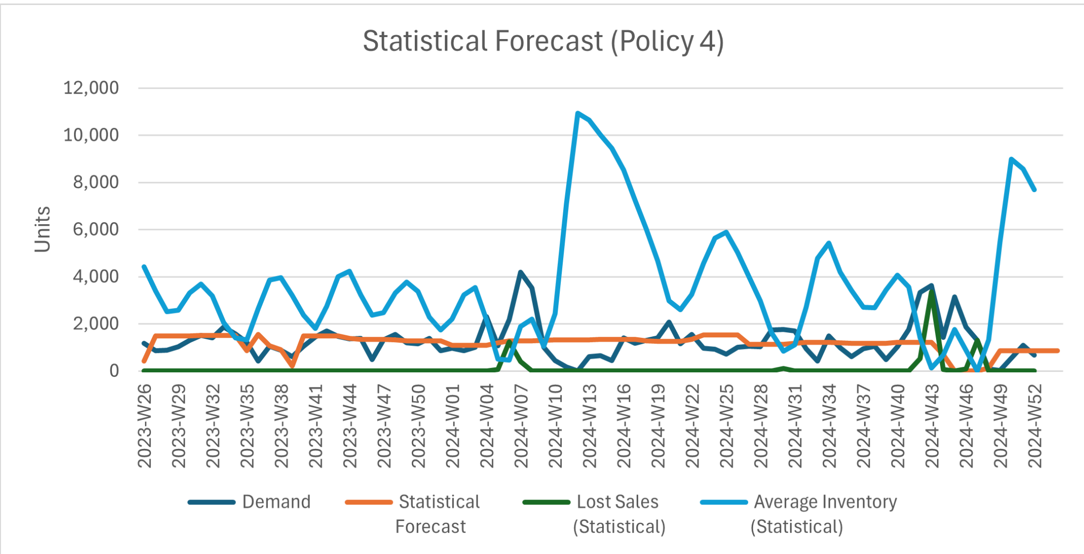

Figure 11: Evolution of Key Metrics - Statistical Forecast (Policy 4)

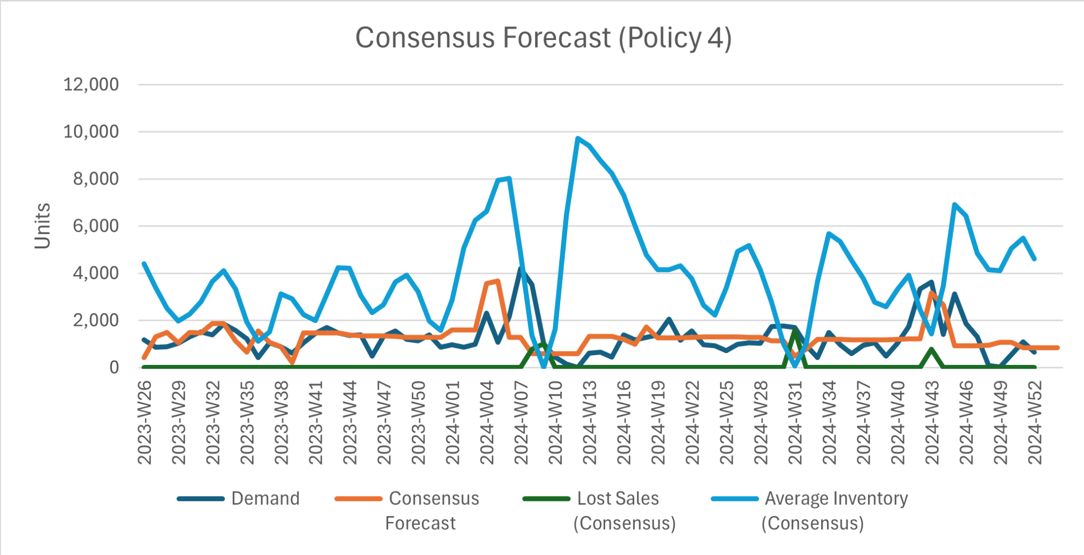

Figure 12: Evolution of Key Metrics - Consensus Forecast (Policy 4)

Policy 4 introduces a sophisticated safety stock model, but its performance raises some concerns. On the one hand, is more reactive to demand fluctuations – safety stock adapts dynamically to forecast accuracy and it has a better risk management – higher inventory prevents stockouts more effectively. On the other hand, we see excessive inventory spikes – at times, inventory surges to over 10,000 units, leading to storage constraints and higher holding costs. Moreover, it is inefficient compared to Policy 3 – this method required more stock for the same service level, making it less optimal in terms of inventory efficiency.

Closing Thoughts

While this policy provides a more data-driven approach, the results suggest that Policy 3 was actually the best compromise. The historical variability method is useful, but relying too much on past forecast errors can lead to overcompensation, causing inventory levels to swing dramatically.

Conclusion

When it comes to forecasting, there's a clear winner: consensus forecast. Not only does it consistently outperform statistical forecasting in terms of accuracy, but it also leads to better inventory management.

As we mentioned at the beginning, no replenishment strategy can be judged in isolation—financial impact is key. My recommendation would be to test all four policies in a real-world setting and evaluate their performance in economic terms.

That said, if we had to choose the best replenishment strategy purely based on our results, Policy 3 (Fully Dynamic Replenishment) stands out. It strikes the best balance between inventory efficiency and fill rate performance. In contrast, Policy 4, while conceptually appealing, is overly reactive, causing stock fluctuations and excessive inventory levels. The first two policies have some merit, but their potential is more limited.

Policy 3 not only delivers strong results now, but with more refined forecasting, it could become an even more robust and scalable solution.