Supply Chains and the Triple Bottom Line

The Problem Today

Supply chains are critical to our everyday life, and we rely on them to provide us with our most basic needs. However, there are two significant challenges that supply chains face that make them particularly vulnerable: (1) high operational costs and (2) increasing pressure to maintain an environmentally friendly supply chain. The second challenge compounds the first given that maintaining a sustainable and ethical supply chain increases operational costs. However, supply chains need to comply with government regulations, and maintain a favorable brand image for customer relevance.

Operational inefficiencies and waste are main contributors to the first challenge, and minimizing deforestation pertains to the second challenge. Deforestation occurs as a result of consumer demands, with agricultural demands playing a significant role (Shutay, July 2022). Deforestation compounds negative effects on the environment (and people) by destroying habitats and ecosystems, and by removing trees that once served as carbon sinks. Furthermore, 14.5% of global anthropogenic greenhouse gas emissions result from livestock supply chains alone[1] (FAO, 2017). Therefore, food production is a major contributor to the problem.

Despite the harmful impacts of food production on the environment and the negative impact of food waste on profit margins, as much as 38% of food is wasted globally[2]. Therefore finding ways to reduce food waste should be a priority for all food supply chains. As such, the focus of this article is on minimizing waste in food supply chains specifically. With that said, the strategies and approaches discussed in this article apply to all supply chains.

Given the challenges that we face, one might ask how we best minimize waste in our supply chains. In the remainder of this article, I will discuss how to approach this problem from a business and data science perspective, and then I will close out with a practical example of how one might select a forecasting approach where the goals include minimizing waste and reducing costs.

---

[1] For a deeper dive into the effects of climate change on our food systems and supply chains, refer to (Shutay, November 2020).

[2] This number was found via a Google search where estimates ranged from 33% to 38%. The source referenced is https://www.rts.com/resources/guides/food-waste-america/.

Selecting a Forecast Approach to Support Supply Planning

Getting an accurate view of demand starts with both inventory management and sufficient sales history. If we do not have a reliable and valid view of inventory and sales, it will be very difficult to develop meaningful demand and supply plans. One of the first steps in the supply planning process is to develop a forecast of future sales.

Developing forecasts to support supply plans is both an art and a science. I say this because we need to rely on subject matter expertise to understand the products that we are forecasting and other contextual factors (or we need to get input from the business), and we need to leverage data science and technology to develop our forecasts. Our approach will depend on several factors. For example, we need to consider the level of granularity to use (time and space), amount of history to include, taking a top-down versus bottoms-up approach, which algorithm(s) to select, as well as which data inputs (features) to include and how best to engineer them.

Here are some of the critical questions that we need to ask to help determine our approach. The first six questions primarily influence the demand plan while the last four questions pertain to the supply plan strategy.

- Is this a new product introduction?
- Is the product heavily promoted (e.g., price, advertising)
- Is the product highly seasonal?
- Is the product slow moving or have intermittent sales?
- Are the sales highly stochastic?
- How much usable history do we have?
- Does the product have a short shelf-life?
- Are inventory levels severely constrained due to holding costs or available storage space?
- Are the case pack size options aligned with demand?
- Are we working with a multi-tiered complex supplier network?

At a very basic level, forecasts are comprised of three components: (1) baseline sales[3], (2) promotion related impacts, and (3) external factors. Promotions can have a negative or positive impact on sales in terms of promotion lift (on promoted items), halo, and cannibalization. When products are new and have no historical sales, or when they are continuously promoted, true baselines do not exist.

[3] Baseline sales in this case are defined as trend plus seasonality with all promotional activity and external factors such as weather or competitor impacts removed.

In the case of new product introductions, the best approach is to identify proxy products to use to gauge the sales for the new product and incorporate promotion-related adjustments to the projected sales. In an ideal world, there will have been a market test to provide a starting point. With continuously promoted items, we tend to work with quasi-baselines to estimate the relative impact of historical promotions. For example, we might decompose the historical sales into a trend, seasonality, and residuals component. The trend plus seasonality comprises the quasi-baseline, and then we can examine the relative impact of the various promotions on our target variable, which in this case would be the residuals.

The visual in Figure 1 shows the output of a decomposition model where the residuals (the deviations from the quasi-baseline) become our target variable. We want to know if promotion-related factors or external factors (e.g., extreme weather, nearby events) explain those deviations and to what extent. We can then use those estimates (e.g., degree of impact) to predict future impacts to sales, and we can scenario plan based on different assumptions and scenarios.

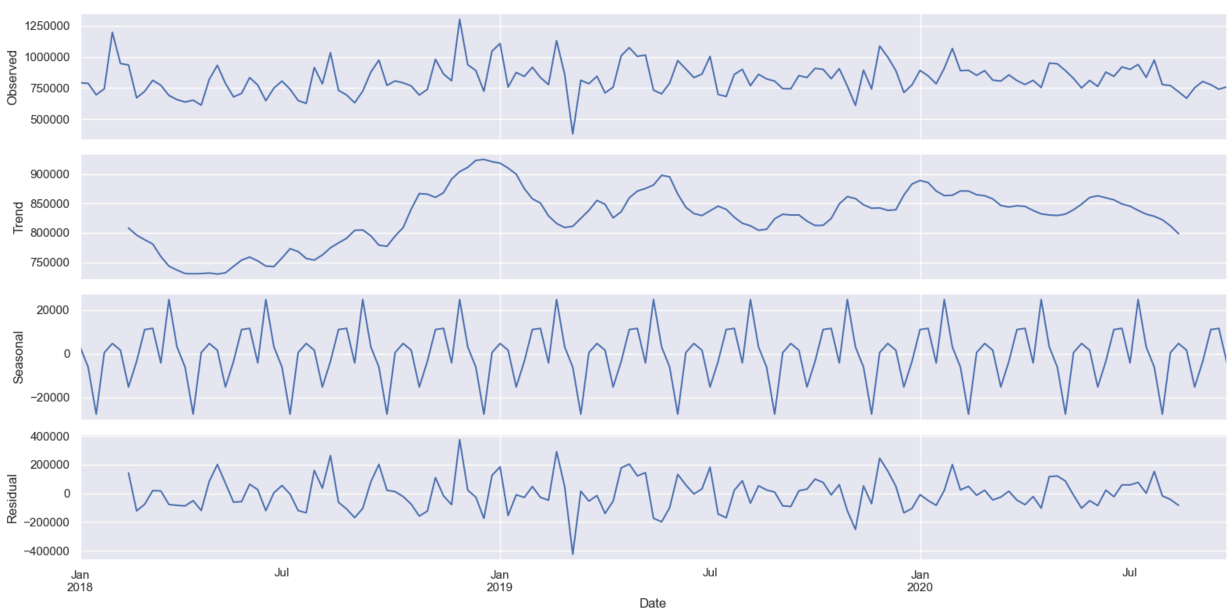

Figure 1. Example Output from a Seasonal Decomposition Model

For many items, there is typically sufficient history to model the actual sales, and we can assume baselines exist. Please note that I am differentiating between baseline and benchmark. A baseline model, as defined in this article, is a series with all promotional activity and exogenous factors (e.g., extreme weather events) removed from the series. A benchmark model is a starting point and is used as our comparison model when iterating for improved accuracy. It is important to note that while the algorithm or model that one selects is important and has implications for forecast and prediction accuracy[4], the data that feeds the algorithm is the most critical component.

There are various algorithms and approaches that one can leverage to generate demand forecasts. Historically, state-space models (e.g., exponential smoothing, autoregression, ARIMA) have been used for time series forecasting. More recently, machine learning and deep learning models have been used and have been shown to provide superior performance (lower forecast error) in many cases[5]. However, forecast error is only one of many factors to consider when selecting an approach.

Another thing to consider when selecting an approach is the benefit of using something simple. Not only are simpler models easier to interpret and manage, but they also generate fewer carbon emissions than complex deep learning models[6]. Therefore, I recommend only using deep learning models when they demonstrate clear benefits. Even in cases where you might gain a slight increase in accuracy, if it costs the business more to get to that level of accuracy than it saves the business and/or the increase in accuracy does not result in a better supply plan, then it is not worth it.

Similar to the concept of a simpler model is the concept of using a single model architecture. There are benefits to using one model architecture, which include (1) having a simplified workflow making maintenance and debugging easier, (2) use of consistent performance metrics making model monitoring simpler, (3) lower computational costs and faster model training and deploying, (4) easier integration into existing workflows and systems, (5) uniformity in predictions so that forecasts do not have large swings from one week to the next (these can be difficult to explain and defend to clients), and (6) a lower likelihood of overfitting.

Finally, it is important to consider how the forecast translates to a supply plan, which ultimately determines the degree of waste and potential lost sales. We can then compute ROI calculations to assess the degree of impact associated with a given forecasting approach.

---

[4] Please note that when I refer to forecast accuracy, I'm referring to forecasted sales specifically. When I refer to prediction accuracy, I'm referring to the ability to predict the impact of an event on sales.

[5] Makridakis, S., Spiliotis, E., Assimakopoulos, V., Semenoglou, A. A., Mulder, G., & Nikolopoulos, K. (2022). Statistical, machine learning and deep learning forecasting methods: Comparisons and ways forward. *Journal of the Operational Research Society*, *74*(3), 840–859. https://doi.org/10.1080/01605682.2022.2118629

[6] It is best practice to use data centers that are associated with low carbon intensity (e.g., renewable energy). DeepLearning.ai offers an excellent short course on how to do this in the Cloud.

Practical Demand and Supply Planning Example

Before embarking on a forecast improvement project, the business should determine the potential economic benefits of the forecast improvement. For example, we should have a sense of what the economic return would be if we improved our forecast accuracy by 5%, 10%, etc. We could do this via simulations and then translate the results to an economic return. We should have a target that aligns with a desired level of waste and cost reduction[7]. If we need to achieve a 5% improvement in our forecast accuracy to reach a net positive profit based on our simulations, then that should be our minimal target. When considering economic return, we also need to consider development and deployment costs as well as any incremental operating costs.

Once we have a sense for what we need to achieve from a forecast accuracy improvement standpoint, then we need to assess the feasibility of reaching that target. What will it likely take for us to get to that level of accuracy and what are the assumed costs associated with it? A proof of concept (PoC) should be conducted first to prove that we can actually reach that target (level of accuracy). Typically PoCs run for about 12 weeks (six two-week sprints). If the PoC results yield the desired result, then we would conduct an A/B test in the wild to assess the operational feasibility and the economic returns. If the test is successful, we would scale the solution in production.

In order to provide an example of how we would conduct a PoC, I have taken a [Kaggle dataset] containing sales history for food items with varying attributes and modeled them using different approaches for comparison purposes. I have provided 13 week MAPEs for each scenario and their respective standard deviations in Figure 2. Please note that these results are based on out-of-box implementations of the four models without any hyperparameter tuning or incorporation of external data[8].

The results indicate that regional models outperformed a top-down approach (national) in most cases[9]. The results also indicate that the error is greater for stochastic items and for low volume items, which is expected. Finally, the differences in performance between the models was smaller for low variability and high volume items, which tend to

[7] If the goal was to minimize lost sales, then our approach would be to avoid breaks in supply. Therefore, we might prefer a model with an oversupply bias and be more tolerant of stranding.

[8] This approach was taken for simplicity purposes given that the goal of this article was not to test the superiority of ML and deep learning models, but to demonstrate how forecast accuracy is only one of many important factors to consider when selecting a forecast approach.

[9] National models may have performed better if the shares (indices) used were based on the same week of year instead of the overall share for the series. The overall share of the series was used in this study to simplify the analysis. In the real-world, we would want to use a more precise index to compute shares.

be easier to forecast. Based on these results, do we have enough information to move forward with an approach?

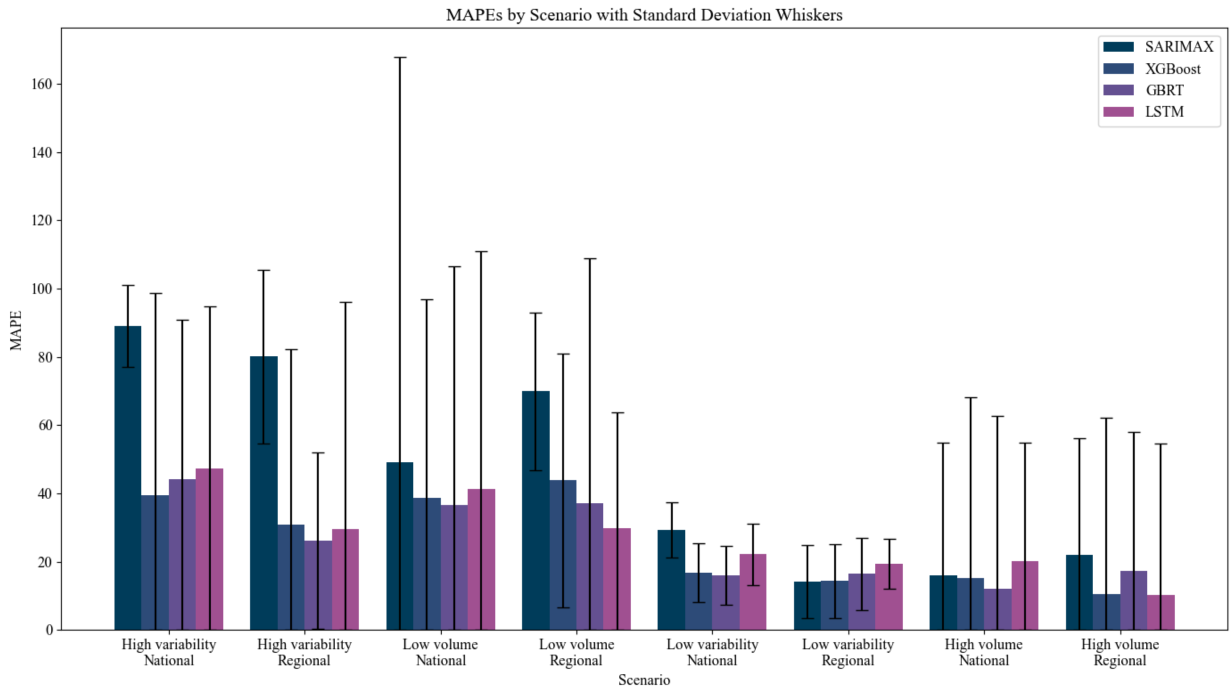

Figure 2. MAPE by Scenario with Error Variability

While these model results are helpful, they don't give us the full picture. We need more information before we can decide which model to put into production. Specifically, we need to know how these forecasts translate to expected waste and profit margins. Therefore, we need to take into consideration the direction of the error (over versus under forecast), the case pack size constraints, shelf-life considerations, and inventory constraints. Once we incorporate these factors into our plan, we can better understand the extent to which our model's error affects waste and profit margins and if the juice is worth the squeeze. Let's do a deeper dive into one of the items from Figure 2 to demonstrate how this process works.

I have taken the high variability item (meal ID = 1971 from the Kaggle dataset) as our use case. Based on the data featured in Figure 3, we can clearly see that the discount variable is the cause of the large spikes in sales in our test set, and the larger the discount, the higher the sales. You will also notice that there is a decrease in actual sales during the discount period, but the forecast assumes more consistent demand in later promoted weeks. There are two possible reasons for this decrease. One reason is a break in supply, which reduces the potential sales for a given week. Another reason is promo decay or fatigue where the target market may become saturated or customers have lost interest in the promotion. Sometimes we have data to confirm this one way or the other, but oftentimes we do not. We can make assumptions when data is not available and adjust the forecast accordingly. This process can be automated for scale.

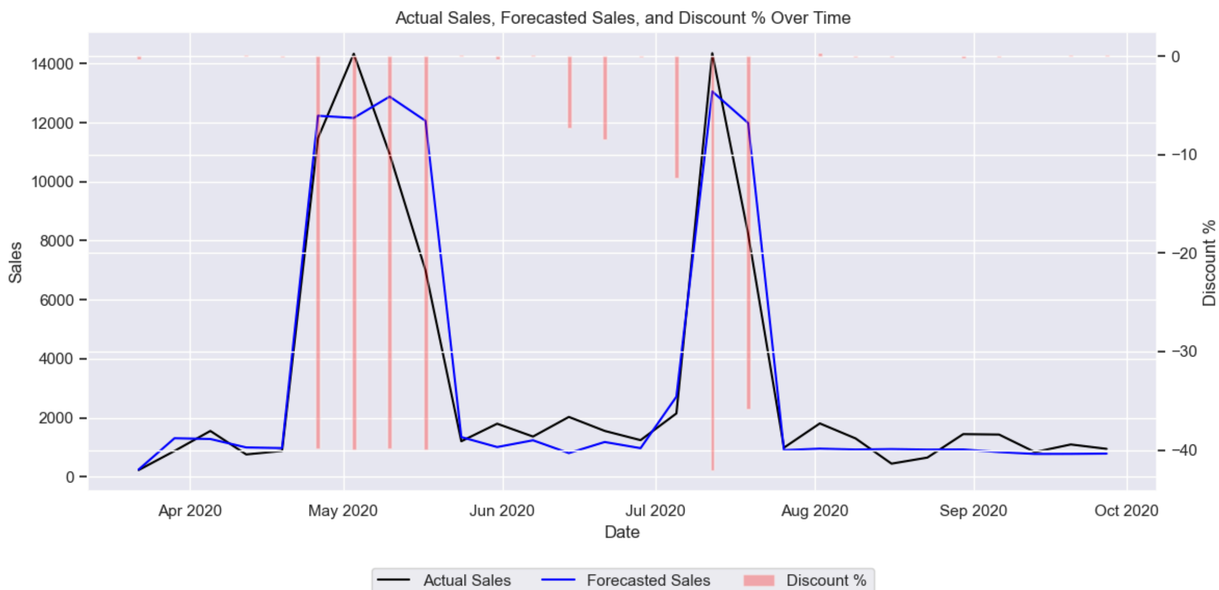

Figure 3. GBRT Model Actuals, Forecast, and Discount Percent

Once you believe that you have your best view of demand based on your best model, you need to translate that to a supply plan. For example, if we have a case pack size of 60 and the demand is 100, we need to decide if we will order one case pack and run short or order two case packs and have 20 units stranded. If we have the ability to hold inventory on hand[10] and the shelf-life permits, we can plan for a two-week supply and carry over stranded products. However, for this use case, I'm assuming that the product expires after one week and therefore any product stranded is wasted[11].

For the sake of this example, let's assume that our XGBoost model is our benchmark model and we want to compare it to our GBRT model that we are considering. Figure 4 provides a comparison between the two models in terms of stranding (positive values) and breaks in supply (negative values) after factoring in case pack size and direction of the forecast error (over versus under supply). Please note that the forecast error for the

[10] If oversupplied products are being stored in inventory, then we need to consider the holding costs as part of our economic analysis.

[11] We should first try to optimize the case pack size to get as close to demand as possible and we should be prepared to offer alternatives to customers when breaks in supply occur, which could help keep customers happy while also reducing waste. For items without a short shelf-life, we can lean towards oversupplying since those items can carry over into future weeks and offset potential lost sales.

GBRT model was 26.2% and the forecast error for the XGBoost model was 31.0%, which is a difference of approximately 5%.

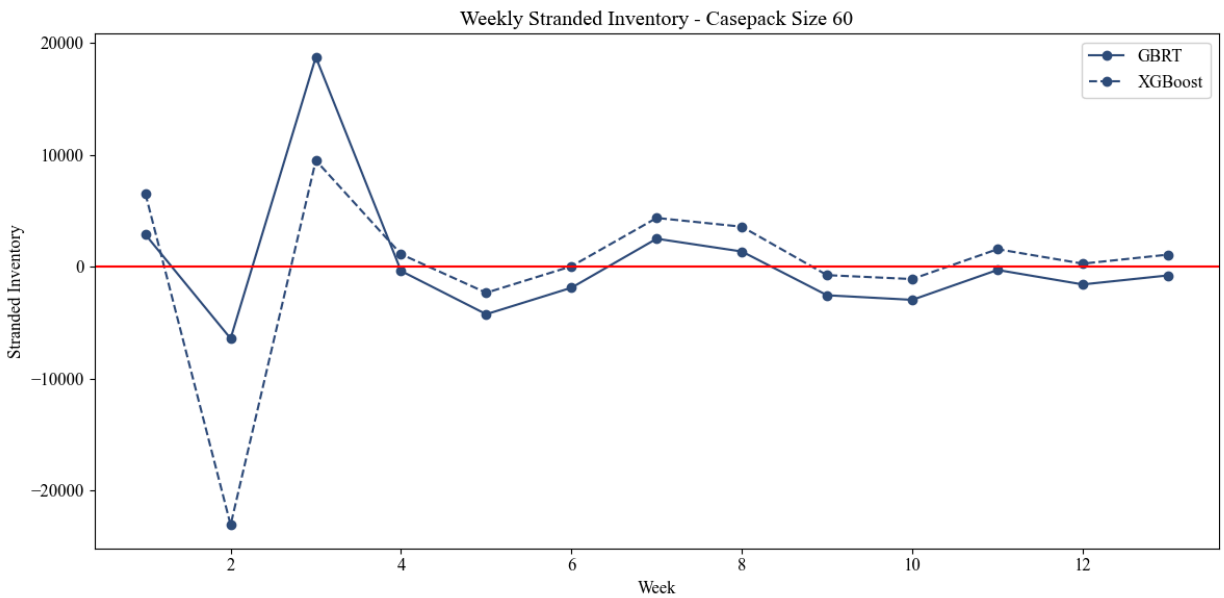

Figure 4. Case Pack Stranding Analysis by Algorithm over 13 Weeks

If we assume that the price of the item is $10[12], the food and paper costs are 30% of the price of the item, and we lose 50% of sales when we have a break in supply, then the difference in profit between the two models over 13 weeks is $155,173 and $122,933, respectively. From a waste perspective, we have 25,515 units wasted versus 28,125. Therefore, if we replace the XGBoost model with the GBRT model, we expect to increase profit by 26% and reduce waste by 9%[13]. Again, we would need to consider development costs and any incremental operating costs to determine if this is a net positive. If so, we would move to the next step and pilot this in the wild and then scale it in production pending a successful pilot.

Even when doing everything right, we will still end up with an oversupply. When this happens, running promotions on oversupplied products is a common and effective strategy. And, this is where Integrated Business Planning becomes incredibly valuable!

---

[12] This amount needs to be adjusted based on the price discount when the product is discounted, but the cost should be based on the full price (30% of full consumer price = cost per item to the business).

[13] We cannot expect this same result in all cases. The actual ROI will depend on over versus under forecasting and how that coincides with price discounts, etc. A 5% forecast improvement will have a wide range of potential ROI results, and therefore a PoC should always be conducted before spending a significant amount of time and resources on implementing at scale.

Conclusions

We are at a point in time where supply chains must place a greater emphasis on the triple bottom line (people, planet, profit). Making supply chains more environmentally friendly will undoubtedly add costs to the chain, but there are opportunities to offset those costs through optimized planning and risk mitigation.

One of the most powerful ways to reduce the negative effects of supply chains on the environment, and to simultaneously reduce costs, is to optimize the planning process to minimize waste. We can do this by getting a more accurate demand signal. However, picking an approach based solely on forecast error could lead you to a situation of diminishing returns or even result in a worse scenario. We also need to consider how that demand translates to a supply plan and ensure that our case pack sizes are optimal. Other economic factors to consider include operational costs (development, deployment, model maintenance, compute, etc.).

The optimal approach will be dependent upon the attributes of the items and the unique strategies associated with the organization of interest. With that said, one should consider the trade-offs of simplifying the approach. Also, it is important to focus on the data that are feeding the model; this includes raw data and feature engineered data. I recommend doing PoCs with external data sources to evaluate the economic returns associated with purchasing such data.

Once we have our approach selected, the goal is to automate as much as possible so that our process can be scaled. Everything that I discussed in this article can be automated.

Note about the Author

Dr. Jeanette Shutay is a strategic research and data science leader with extensive experience building, scaling, and leading advanced analytics and data science teams. Jeanette has over 20 years of experience leading innovative and applied research projects and over 10 years of consulting experience focusing on data strategy, research methodology, program and campaign evaluation, data science, and decision sciences. Currently, Jeanette is the President and Chief Data Officer at Shutay Consulting, and she is a Lecturer in the MS in Applied Data Science program at the University of Chicago.