Outline for IJF paper on Whether Forecast Accuracy Matter Approximately 15 pages

**Title:**

Does Forecast Accuracy Matter: The Impact of Accuracy Improvement on Supply Chain Outcomes

**Article Keywords:** Forecast Accuracy, Supply Chain Outcomes, Business Application, Accuracy Relationship, Demand Planning

**Abstract:** Consulting and commercial organizations have frequently referenced the business and supply chain benefits of increasing forecast accuracy. The question of the relationship between forecast accuracy and business benefits is important for forecasting practitioners and researchers. Deciding whether to attempt to implement new forecasting models or modify parameters of existing models involves significant effort. By identifying the impact of improving forecast accuracy on supply chain outcomes, forecasters can apply a benefits lens to prioritize their efforts and determine the expected returns from their efforts. In this paper, we propose an approach for identifying the business benefits by utilizing a supply chain modeling method and perturbating forecast predications with directed error levels to identify the relationship between forecast accuracy and key supply chain output measures.

**Introduction**

Forecasting demand, particularly to support supply chain decisions, is one of the most frequently used implementations of forecasting in business. While there have been numerous papers, both commercial and academic, that compare forecast accuracy among specific forecasting methods, there are very few published research papers that explain how improving forecast accuracy, without regard to forecasting method, affects supply chain outcomes. Those papers that have been published normally compare specific methods (e.g., Naïve forecast versus exponential smoothing or machine learning methods). The range of differences in forecasting accuracy between methods is normally relatively small. Small differences in accuracy in these papers have then been utilized to measure the change in supply chain outcome. These comparisons in outcomes are developed using simulations of the decisions that are implemented in specific supply chain algorithms. Because the differences in accuracy are small, it is difficult to generalize the results of the supply chain outcomes across a broader range of forecast accuracy levels. One of the key research questions in this paper is, “As you improve forecast accuracy, are the impacts on supply chain outcomes consistent, or are there points where increased accuracy fails to yield the same level of supply chain improvement?”

Answering this question has significant implications for the process of trying to improve your forecast accuracy. In large firms making thousands, or millions of demand forecasts per day or month, determining which forecasts should be worked on for improvement is an important process question. Firms do not have unlimited resources to apply to this effort (human, software, algorithms, or hardware). This research, by applying consistent levels of accuracy (error), to the actual demand value, permits the supply chain outcome assessment for each Stock Keeping Unit (SKU) to be measured without regard to a forecast method. Doing so permits supply chain managers and forecasters to estimate the impact of improving the forecast accuracy for each SKU on key supply chain outcomes. This information can then be used to prioritize forecast improvement efforts and estimate the effect that a given level of forecast accuracy improvement will have on key supply chain outcomes.

**Material and methods**

In order to make this research replicable, the dataset used to represent demand values to be forecast was the daily demand of 30,490 SKUs provided by Walmart to support the MOFC, M5 Forecasting competition (<https://mofc.unic.ac.cy/m5-competition/>). The dataset of daily SKU demand can be retrieved from the Kaggle.com competition website (<https://www.kaggle.com/competitions/m5-forecasting-accuracy>). The file is sales\_train\_validation.csv. This file contains information about each SKU followed by 1913 daily observations of demand for each SKU from 2011-01-29 through 2016-04-04.

Forecasts with known levels of accuracy were developed at these levels of error: 50%, 40%, 30%, 20%, 10%, 5%, 2%, 1%, and 0% (i.e., no forecast error). These forecast errors were introduced by utilizing the actual demand for each SKU for each period of time and calculating the error (either plus or minus) and adding (or subtracting) the error from the actual value of demand at each of the different error percentages. This created nine different forecast values per day for each SKU.

In addition to the forecasts at the different known levels of accuracy (error), it is necessary to adopt a supply chain simulation method to capture the impacts of the different levels of accuracy. One of the most popular implementations of supply chain software in the world can be found in software provided by SAP Supply Chain Management (SCM). SAP’s software has been implemented by over 8,229 companies globally (enlyft, 2023). All of the calculations and input variables necessary for the supply chain outcome calculations were derived from the documentation for that software (SAP, 2021).

To estimate the supply chain outcomes for each SKU at each level of forecast accuracy (error), a simulation was run. The code is designed to permit any number of SKUs to be estimated at each level. The simulation code was developed in the data science program R, and utilized RStudio as the IDE. The code is available at: <https://github.com/Hoover-code/Forecasts_matter>. The file name is demand-v23.R. Supply chain outcome results are output in Microsoft Excel format which can then be used to assess results across SKUs.

**Calculation Methods**

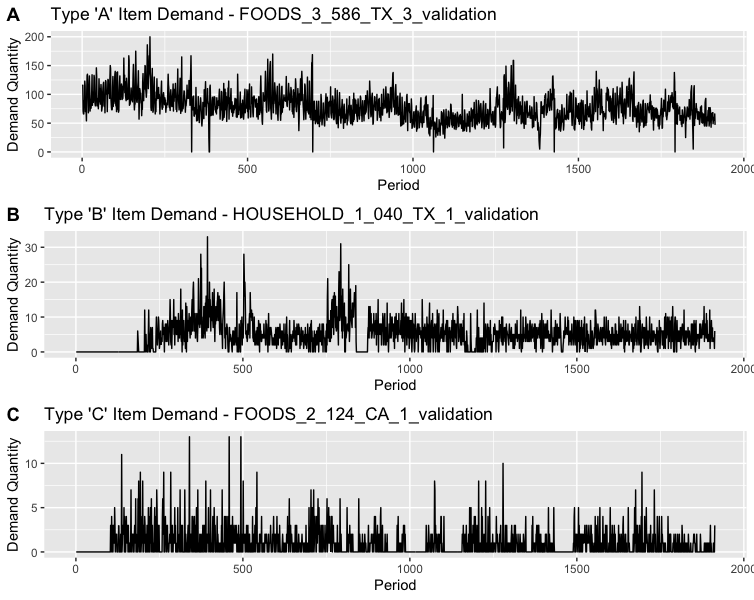
Kolassa (2022) and Robette (2022) proposed utilizing a simulation framework to assess the impact of improved forecast accuracy on supply chain outcomes. Neither proposed utilizing multiple known levels of forecast accuracy (error) within the simulation to assess the impact of improving the forecast on supply chain performance. The simulation approach for the assessment followed the steps below:

1. Decide on the key supply chain performance outcome measures. In this case, we utilized the following measures at each level of forecast accuracy:
   1. average inventory on hand
   2. backorder quantity per period
   3. percentage of total periods where backorders occurred
   4. backorder quantity / demand (i.e., fill rate)
   5. inventory holding cost
2. Collect time series and other data for the simulation. I this case, I utilized the SKU demand data from the M5 competition as the actual demand per period. Other fields are required to perform the supply chain simulation. These are not found in M5 competition data, and were assumed for simulation purposes. Some key fields within the simulation that were assumed include:
   1. Reorder Point, Fixed Order Quantity (R,Q) inventory replenishment policy
   2. Replenishment lead time (in days) – assumed fixed
   3. Forecast period (in days) – assumed to match replenishment lead time (in days)
   4. Ordering cost per order
   5. Holding cost per item
   6. Service level
3. Simulate multiple forecasts for each time period. In this step, forecasts were calculated using known intervals of forecast error. Those errors were either subtracted or added to the actual value that day to achieve a forecast. Below is an example calculation of the error and subsequent forecast when the actual value was 52.  
     
   A screenshot of a computer error

   Description automatically generated

Figure 1- Example calculation of forecast at different levels of forecast error.

1. Following the creation of the forecasts at the nine different levels of forecast accuracy (forecast error), a simulation of the supply chain decisions was created utilizing those forecasts. The simulation utilized the calculation methods employed by one of the most popular enterprise level supply chain planning systems.
2. Supply chain outcome metrics were captured for each SKU at each simulated level of accuracy. The outcome metrics included: a) average inventory on hand; b) number of periods with backorders; c) backorders as a percentage of total demand; total backorder quantity; d) fill rate; e) total number of replenishment orders; f) safety stock average during the simulation; g) inventory holding cost; and h) average daily demand.
3. To normalize the outcome metrics to compare across SKU’s the following additional supply chain outcomes for each SKU at each level of accuracy were calculated: a) days demand on hand; b) safety stock days of demand; c) inventory holding costs / days of demand; d) orders to average demand ratio; e) reorder point to average demand ratio; f) backorders to average daily demand ratio.
4. A metric to classify the SKU demand was added to the data. The 30,490 SKUs in the Walmart data illustrated the following distribution of average daily demand.  
     
   A graph of a number of daily demand

   Description automatically generated with medium confidence  
     
   This average daily demand pattern reflects a typical pattern found in categorizing demand using ABC analysis (Dickie 1951). The A, B, C levels for the average daily demand for the Walmart SKUs was classified based on the following algorithm: 1) “A” where demand was greater than or equal to 10 demands per day (N = 365); 2) “B” where demand was greater than or equal to 2 demands per day, but less than 10 demands per day (N = 3,507); 3) “C” where demand is less than 2 demands per day (N = 26,618). These classifications are especially important to the analysis of supply chain results since items in category C will have many daily periods where the demand observed is 0. The forecast using this paper’s defined forecast error rate calculation method will also be 0 in those periods. Below is a graphical representative SKU example of demand for items in each of the 3 categories. Note that many periods have a demand value of 0 in the C category item.  
     
   

Simulating the forecasting process and calculating the outcome metrics at each level of forecast error took approximately 1 minute per SKU. Blocks of 1,000 individual SKUs were distributed across multiple computers to calculate the outcomes and store the results for analysis.

**Results**

Traditionally, forecasters have strived to increase forecast accuracy with a goal of having no forecast error (or 100% forecast accuracy). Additionally, based on anecdotal and survey research, managers expectations of forecast accuracy are quite high (Morris, 2014) (Tashman & Hoover, 2020) in the 90% to 95% accuracy range. A 2019 Forecasting and Inventory Benchmark Study by the website Supply Chain Brain found that this high expectation of forecast accuracy by management is driven by the aggregated financial projection forecasts that they produce for quarterly reports (E2open, 2019). This study indicates that forecast accuracy varies by business decision level in the supply chain. At the weekly – item – location level which aligns with the SKU data provided by Walmart, the average forecast accuracy achieved is about 52%. Average accuracy was lower during the pandemic period at 45%.

Garner consulting group (Steutermann, 2017) indicated that improving forecast accuracy by 1% can result in: 1) a 2.7% inventory reduction; 2) a 3.2% transportation cost reduction; and 3) a 3.9% inventory obsolescence reduction.

While these two studies indicate that there is room for improvement in forecast accuracy and that those improvements impact supply chain performance, others including (Robette, 2023) and (Kolassa, 2023) indicate that the performance improvements aren’t always as straightforward as presented.

In this study, the results of the supply chain outcomes at different levels confirm those cautions. By analyzing the results of supply chain outcome metrics at different levels of forecast accuracy for each SKU and at each ABC level, we can see some findings that run counter to the improvement expectations assumed. Additionally, the improvements are not linear as forecast accuracy improves.

**Ratio of Safety Stock to Average Daily Demand**

Safety stock is a small portion of the total inventory cycle relative to the average daily demand. Inventory levels are dominated by the reorder point or the reorder quantity given the replenishment cycle and the economic order quantity. Safety stock accounts for the variance of demand over lead time. As such, we found that the ratio of safety stock to average daily demand varies from a high of 1.2 days of demand and drops to near 0 days of demand as you reduce forecast error from 50% to 10%. Beyond a forecast error of 10% (i.e., 90% accuracy, the days of safety stock drops to 0 and provides no coverage for the variance in daily demand.

A graph of a graph with red dots

Description automatically generated with medium confidence

**Backorder Periods as a Percentage of Total Periods**

Because safety stock levels decrease to zero as forecast accuracy improves, the number of backorders will increase as the remaining inventory levels created by the reorder point calculations must absorb the entire variance of daily demand. This effect is especially true for the higher demand category (the “A” demand classified items). In the A category items, the percentage of periods (days) with backorders begins at 1.3% when forecast error is 50% and drops below 1% as accuracy improves to 80%. But, the number of periods with backorders increase as forecast accuracy increases to 90% or better. As forecast accuracy increases to 100%, the A category backorder periods increases to 1.35% (even higher than when forecast accuracy was 50%). For the low demand (Category C) items, this effect does not occur because the reorder quantity is much more likely to cover the low variance in demand for these items.

A graph of a graph with a line

Description automatically generated with medium confidence

**Fill Rate Percentage**

The fill rate also reflects the effect of the impact of reduced safety stock as forecast accuracy improves. The highest accuracy in both the categories A and B items is highest at the 80% accuracy level. Beyond that point, the improved forecast accuracy reduces the fill rate for these two categories as the calculated safety stock levels drop to 0.

A graph of the number of percent error

Description automatically generated with medium confidence

**Inventory Holding Costs**

Because inventory holding costs are driven by the average inventory cycle (reorder quantity and safety levels), they display the expected reduction as forecast accuracy improves. Mean inventory holding costs are reduced as the volume of daily demand is smaller and reorder quantities are likewise smaller.

A graph with numbers and a line

Description automatically generated

While the inventory on hand balance is reduced as forecast accuracy is improved, an interesting and important observation is that the overall reduction in inventory holding costs is relatively small. For the faster moving Category A items, the inventory holding cost reduction is 3.9% as forecast accuracy improves from 50% to perfect (100% accuracy). This relatively small improvement is because most of the inventory balance is driven by the reorder quantity and relatively little is caused by the safety stock. For the slower moving item categories, this inventory holding cost improvement from 50% accuracy to perfect is even smaller: Category B items: 2.1% and Category C items: 1.3%. This implies that even dramatic improvements in forecast accuracy have limited effect on inventory holding costs. The ratio of safety stock to the reorder quantity and variance of demand in the item will determine the size of the potential impact.

Discussion

This should explore the significance of the results of the work, not repeat them. A combined Results and Discussion section is often appropriate. Avoid extensive citations and discussion of published literature.

Conclusions

The main conclusions of the study may be presented in a short Conclusions section, which may stand alone or form a subsection of a Discussion or Results and Discussion section.

Appendices

If mathematical derivations are needed, put them in an Appendix.

Dickie, H. F. (1951). ABC inventory analysis shoots for dollars not pennies. *Factory Management and Maintenance, 109*(7), 92-94.

Morris, M. (2014, January 29). Managing Executive Level Expectations of Forecast Accuracy: Yamaha’s Experience. *Institute of Business Forecasting & Planning.* <https://demand-planning.com/2014/01/29/managing-executive-level-expectations-of-forecast-accuracy-yamahas-experience/>

Tashman, L. and Hoover, J. (2020).

E2open. (2019). 2019 Forecasting and Inventory Benchmark Study [White paper]. *Supply Chain Brain*. <https://www.supplychainbrain.com/ext/resources/0-whitepapers/E2Open/E2open_2019_Forecasting_and_Inventory_Benchmark_Study_White_Paper1.pdf>

Steutermann, S. (2017). Win the Business Case for Investment to Improve Forecast Accuracy. Gartner. <https://www.gartner.com/en/documents/3701817>

Kolassa, R. (2023). Using Simulation to Determine When Forecast Accuracy Matters. *Foresight:* *The International Journal of Applied Forecasting.* 2023Q1, Issue 68, p20-24.

Robette, J. (2023). Does Improved Forecast Accuracy Translate to Business Value? *Foresight: The International Journal of Applied Forecasting,* 2023Q1, Issue 68, p12-19.