

# Forecasting method selection in a global supply chain

Everette S. Gardner, Jr.<sup>\*</sup>

Department of Decision and Information Sciences  
C.T. Bauer College of Business  
334 Melcher Hall, Houston, Texas USA 77204-6021  
Telephone 713 743 4744  
[egardner@uh.edu](mailto:egardner@uh.edu)

Yavuz Acar

Department of Management, Bogazici University  
34332 Bebek, Istanbul, Turkey  
Telephone 90 212 359 7380  
[yavuz.acar@boun.edu.tr](mailto:yavuz.acar@boun.edu.tr)

---

<sup>\*</sup> Corresponding author

Jan. 26, 2010

# **Forecasting method selection in a global supply chain**

# **Forecasting method selection in a global supply chain**

## **Abstract**

In supply chains, forecasting is an important determinant of operational performance, although there have been few studies that select forecasting methods on that basis. This paper is a case study of forecasting method selection for a global manufacturer of lubricants and fuel additives, products usually classified as specialty chemicals. We model the supply chain using actual demand data and both optimization and simulation techniques. The optimization, a mixed integer program, depends on demand forecasts to develop production, inventory, and transportation plans that minimize total supply chain cost. Tradeoff curves between total cost and customer service are used to compare exponential smoothing methods. The damped trend method produces the best tradeoffs.

## **Key words**

Demand forecasting, exponential smoothing, inventory forecasting, model selection, simulation, supply chain

# Forecasting method selection in a global supply chain

## Introduction

Several recent studies have found little relationship between average forecast accuracy and inventory performance. Tiacci and Saetta (2009) compared two simple forecasting methods in a simulation study of a warehousing operation. The methods produced similar total costs, but the least accurate method gave better customer service. In another warehousing simulation, Sanders and Graman (2009) showed that bias in the forecasts was more important than average accuracy in determining warehousing and inventory costs. Using real data drawn from a distribution inventory, Syntetos et al. (2009) showed that marginal improvements in average accuracy led to much larger improvements in inventory costs.

This paper examines the relationship between forecasting and operational performance in the broader context of production and distribution within a supply chain. Previous research has established that forecasting determines the value of information sharing (Zhao et al., 2002), and forecast errors contribute to the bullwhip effect, the tendency of orders to increase in variability as one moves up a supply chain (Chandra and Grabis, 2005; Dejonckheere et al., 2003, 2004; Zhang, 2004). Therefore it makes sense to select forecasting methods in supply chains according to the impact on operational performance, although this has been done in only a few studies, as discussed in Gardner (2006). The lack of research in forecasting for supply chains is perhaps understandable because of the expense; usually, one must build a comprehensive model of the supply chain in order to project performance.

This paper presents such a model for the supply chain of a global manufacturer of lubricants and fuel additives, products usually classified as specialty chemicals. The model includes four manufacturing plants and daily time series of actual demand collected over a four-year period. Both optimization and simulation techniques are used to develop production schedules, inventory targets, and transportation plans. Optimization depends on demand forecasts supplied by exponential smoothing and is done with a mixed integer program to minimize total variable supply chain costs.

Management asked for forecasting methods that were simple and easily automated, making some form of exponential smoothing the only reasonable choice. We considered three methods: simple exponential smoothing (SES), Holt's additive trend (Holt, 2004), and the damped additive trend (Gardner and McKenzie, 1985). SES and the damped trend are obvious choices based upon their long record of success in empirical studies (Gardner, 2006). The data suggest that the Holt method would not do well, but it was retained as a benchmark for the other methods. To select the best method, tradeoff curves were computed between total supply chain cost and several measures of customer service. The damped trend gave the best operational performance for any level of cost, followed by SES and Holt. It is interesting to contrast these results to traditional method selection based on forecast accuracy measures; surprisingly, SES gave the best average accuracy.

## **The Supply chain model**

The company produces lubricants and fuel additives for automobiles, farm equipment, marine vessels, trains, construction equipment, and power tool motors. There are four manufacturing plants, in North and South America, Europe, and Asia. Scheduling of production orders is best described as a combination of push and pull processes. Ten component chemicals are produced in a push mode based on forecasts from one to six months ahead, while end products are produced in a pull mode by blending the components according to individual product recipes as customer orders are received. Forecasting is done at the component level by aggregating component requirements across end products at each plant. Considerations of technology, production and shipping costs, and plant capacity are such that not all components are produced in all plants. Thus we have only 25 time series of component demands rather than 40 (4 plants x 10 components).

The supply chain model integrates optimization and simulation and performs tactical planning at two levels. At the first level, the model produces a monthly master production schedule and a stock transfer plan over a six-month planning horizon. These plans are generated by a mixed integer programming (MIP) model that incorporates demand forecasts as described below, pending orders, beginning inventory levels, machine and storage capacity, alternative modes of transportation, and shipments in transit. The model also incorporates company business rules for minimum run lengths and transportation carrier selection. The objective is to minimize total variable supply chain cost, including the costs of production, transportation, inventory carrying, and import tariffs. Tactical planning at the second level uses another MIP model to break

down the first level results into a detailed weekly production schedule for each machine at each plant over a 12-week planning horizon. For a complete mathematical formulation and solution methodology for the MIP models, see Acar et al. (2009a). The models in Acar et al. are developed using simple assumptions about demand, whereas in this paper we study the behavior of the models when they are driven by a forecasting system for real data.

The simulation model at the second level executes the manufacturing plans on a daily basis, using actual daily demand history that occurred over a four-year period. The model reads the first-level production schedule and manufactures components accordingly, placing them in inventory. Production is measured in tons, and total demand for the last year of operations totaled about 250,000 tons. As customer orders arrive, demand is met by blending end products from component inventory. There are a total of 15 machines in the four plants, and production lead-times range from two to seven days, depending on product and order size. If available inventory is not sufficient to meet demand, backorders are generated. The second-level model also transfers stock between plants as required, debiting inventory from the sending plant on the departure date and crediting inventory at the receiving plant on the arrival date. There are numerous transshipments between plants, and the average transportation lead-time is 47 days. All shipment quantities are set as close as possible to those determined in the first-level model (based on inventory availability). If a shipment quantity is significantly less than suggested in the first-level model, another shipment cannot be scheduled until that model is run again.

There are three sources of uncertainty in the simulation. First, actual demand is of course uncertain. Second, transportation lead-times are generated from a set of truncated normal distributions, one for each source-destination pair. Means and standard deviations are based on actual experience, and the distributions are truncated such that the minimum lead-time is 85% of the mean. The reason for the truncation is that most transportation is by marine vessel, and it is impossible to achieve lead-times any shorter. Finally, there is some supply uncertainty due to machine breakdowns. We did not have empirical data to develop distributions of machine breakdowns, and we chose the following simulation procedure in consultation with maintenance and supply chain managers. The occurrence of breakdown for each machine was generated from a uniform distribution; for each breakdown, the duration was generated from a normal distribution with a mean of five days and a standard deviation of two days. It might seem that the probability of breakdown should increase with time, but the company disagreed because rigorous maintenance schedules were enforced. Managers reviewed the simulated breakdowns and found them to be reasonable.

At the end of each week, the second-level model records inventory levels, pending orders, quantities shipped to other plants, and costs incurred. The model also records two measures of customer service: number of orders late and weighted lateness. The latter measure, considered the most important by management, is defined as the number of days an item is backordered times the backorder quantity.

When the second-level model completes the last week of the month, the first-level model is run again. Otherwise, the second-level model calculates production targets for the remainder of the month. The calculations are based on week-ending inventory levels,



pending orders, consumption of the forecast within the current week, and scheduled incoming shipment quantities. Also considered are the end-of-month inventory targets, any shipments scheduled to depart later in the month, and safety stocks. There are many alternative procedures for computing safety stocks in the literature. We chose the bootstrap procedure of Snyder et. al. (2002), which was incorporated in the simulation model and used to compute safety stocks after the forecast updates described below. The Snyder et al. procedure has several advantages. It is easy to implement, tailored to lead-time demand, and does not require normally distributed demand, an assumption that cannot be supported in most inventory systems.

The combined optimization and simulation model was run 300 times: 20 replications times three forecasting methods times five levels of safety stock corresponding to Z-values of 0.5, 1.0, 1.5, 2.0, and 2.5. The precision test of Law and Kelton (2000) showed that 20 replications were sufficient to achieve what we considered to be reasonable 90% confidence limits around each measure of customer service.

## **Forecasting**

At the end of each month, forecasts required in the first-level model, from one to six months ahead, are updated using one of the exponential smoothing methods. Before forecasting, the 25 time series of demand were aggregated from daily to monthly. At first we considered using weekly or biweekly time series, but most series presented intermittency problems. All zero observations disappeared in the monthly series, although many observations are near zero. The monthly series are not homogeneous, and it is difficult to generalize about their properties except to say that they are nonseasonal

and very ill behaved. Some series display a relatively constant level with extreme variance, as in Figure 1, series 1. Some display drastic shifts in level like series 2, while others display changes in variance like series 3. There are trend patterns in more than half the series, and all are erratic as in series 4.

The exponential smoothing methods were fitted to the first three years of data, with the last year reserved as a holdout sample to evaluate performance measures in the simulation model. During the last year, the methods were not refitted, and forecasts were made from one to six steps ahead following monthly updates of method components. The methods could have been fitted to less data, with a longer holdout sample. However, it was not clear that supply chain performance measures stabilized until near the end of the third year of operations, so we waited until the fourth year to evaluate the methods. The smoothing parameters in all methods were selected from the  $[0,1]$  interval by using the Excel Solver to minimize the mean squared error (MSE) over the first three years. SES was initialized with the mean of the observations in the fit periods, while the damped trend and Holt methods were initialized with a linear regression on time.

Different procedures for initialization and fitting were tested, although they made no significant difference in cost and service performance during the last year of operations. For example, to confirm parameter optimization, we restarted the Solver several times from different initial positions in each series, but found little difference in performance. There was also little difference when parameters were optimized simultaneously with initial values of method components, when the methods were fitted to minimize the mean absolute error, and when the methods were initialized using only the first six months or the first year of data.

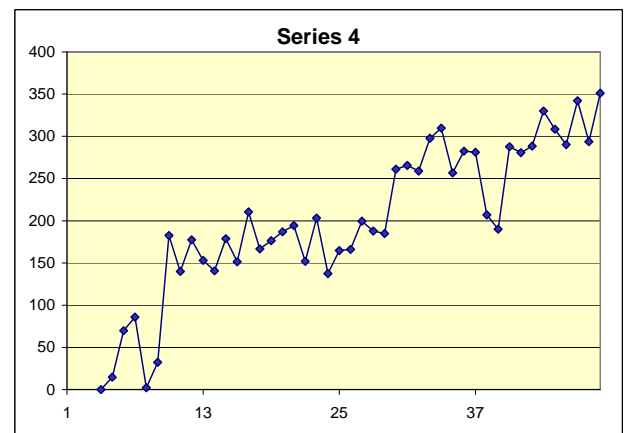
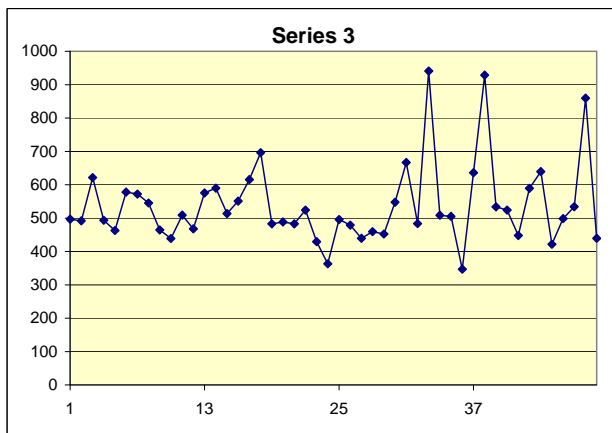
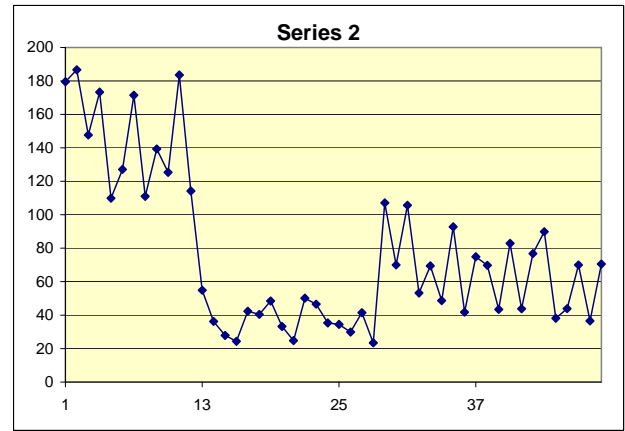
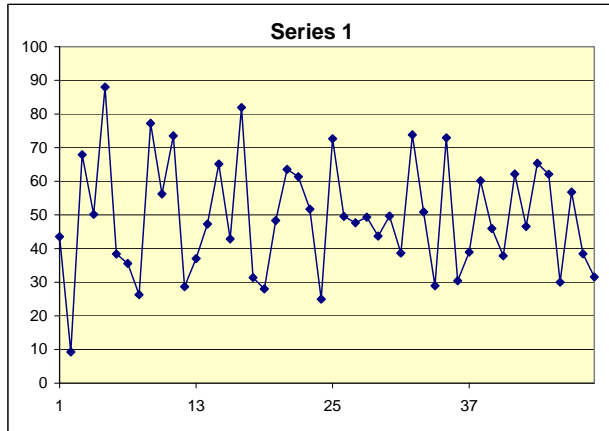


Fig. 1. Examples of monthly time series of end-product demand.

## **Tradeoff analysis**

Gardner (1990) recommended the use of tradeoff curves to evaluate the operational performance of forecasting methods, and we followed that example here. For the last year of operations, Figure 2 gives a tradeoff curve for total supply chain costs vs. average weighted lateness, and Figure 3 gives another curve for costs vs. numbers of backorders. The plotting symbols on each curve represent the five levels of safety stock, and the corresponding costs and service measures are averages of the replications at each level. To put the numbers of backorders into perspective, there were about 41,000 customer orders processed during the last year of operations, and the percentages of backorders were rather large, ranging from about 7% to 19% for the damped trend for example. We were concerned by these percentages, but management felt that the numbers were reasonable in view of capacity constraints. All four plants worked near capacity, both during the last year of the simulation and in actual operations in recent years. Fortunately, most of the backorders were of relatively short duration.

In both Figures 2 and 3, the damped trend gives the best tradeoffs, that is the lowest cost for any customer service level, followed by SES and Holt. For example, management believed that a cost target of about \$115 million was appropriate for this system; at that cost, the damped trend produces weighted lateness of about 118,000 ton-days compared to 134,000 for SES. At the same cost target, the damped trend produces 6,100 backorders compared to 7,600 for SES. For the Holt method, this cost target could not be achieved at reasonable levels of weighted lateness or backorders. In both Figures 2 and 3, as costs and safety stocks increase, the differences between methods become smaller as should be expected, although the differences are always significant.

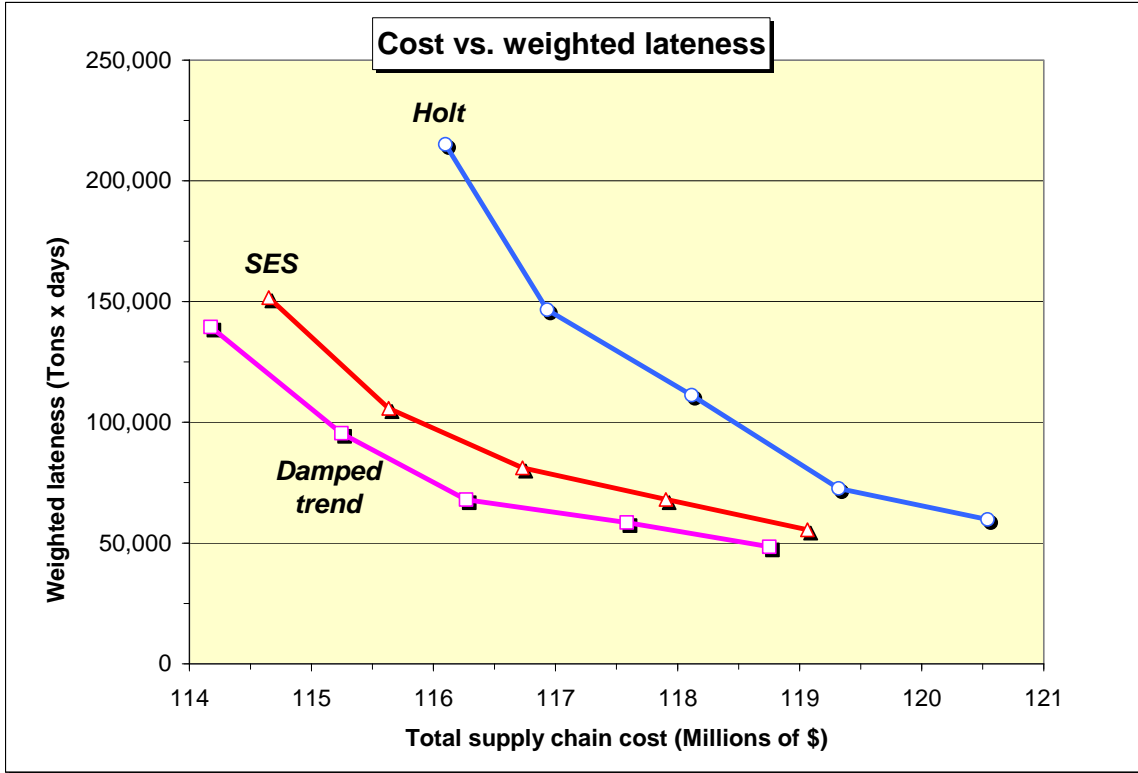


Fig. 2. Tradeoffs between total supply chain cost and weighted lateness (tons backordered times days on backorder) during the last year of operations.

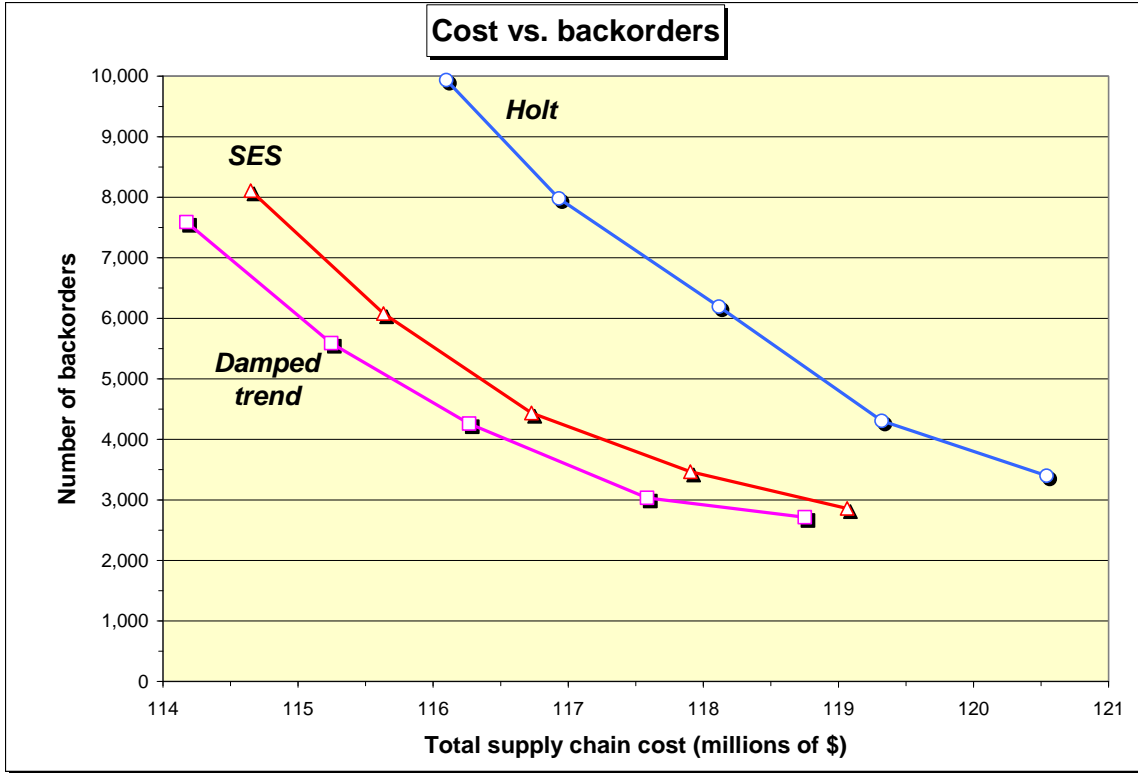


Fig. 3. Tradeoffs between total supply chain cost and number of backorders during the last year of operations.

Production costs account for an average of 90% of total costs for each forecasting method, and we found that these costs do not vary significantly between methods. Management believed that this made sense because all plants operated near capacity. The production process is highly automated, and expediting or overtime related costs are minimal when shortages develop.

The remaining costs are related to transporting and carrying inventory, and they are sensitive to the choice of forecasting method. The consequences of forecast errors are extremely complicated, and there is no simple explanation of why the damped trend produced the best tradeoff curves. To illustrate the problem, consider the effects of under-forecasting for a single product, which of course creates backorders. But backorders usually develop not just for that product, but often for others as well. What happens is that under-forecasting sets off a chain reaction due to capacity constraints. To fill backorders for a single product, capacity is borrowed from routine production schedules for other products, and they in turn can suffer shortages of stock at a later date. When backorders are filled, emergency shipments are necessary, with transportation costs greater than the costs of routine shipments. On the other hand, over-forecasting for a single product means that capacity has been put to the wrong use, and excess stocks are created. What is surprising is that over-forecasting can also create backorders. During the time that limited capacity is devoted to building excess stocks, products competing for that capacity may incur backorders and later emergency shipments as capacity becomes available.

Keep in mind that all 25 products contribute forecast errors that interact with each other in allocating production capacity, and the system is dynamic with monthly updates.

This means that forecast errors are sometimes reversed before the system has fully responded to previous backorders or excess stocks. Given this complexity, the best that can be done in explaining the performance of the damped trend is to say that the method consistently generated fewer backorders and costly emergency shipments than did the other methods. The damped trend did especially well in avoiding backorders for a group of four critical products that account for about 3/4 of total system demand.

### **Average forecast accuracy**

It is interesting to compare the average forecast accuracy of the exponential smoothing methods with their customer service and cost performance. To evaluate accuracy, traditional measures (such as those used in the M- and M3 competitions) are useless in these series because they vary drastically in scale, with some observations near zero. Therefore, we followed the advice of Hyndman and Koehler (2006) and scaled the errors based on the *in-sample*, one-step errors from the naïve method. The mean absolute scaled error (MASE) is thus the mean of the absolute values of the scaled errors, and the mean square scaled error (MSSE) is defined analogously.

Scaled error measures for horizons 1-6 are shown in Table 1. For all products, SES is the most accurate method at every horizon, followed by the damped trend and Holt. However, for the group of critical products mentioned above, the damped trend is the most accurate method. It should be noted that Table 1 is greatly affected by outliers, and removal of a few selected series from the averages for all products would change the rankings.



			Horizon						
			1	2	3	4	5	6	All
MASE	All products	Holt	1.08	1.08	1.11	1.16	1.20	1.28	1.15
		Damped	1.00	0.98	1.02	1.05	1.07	1.11	1.04
		SES	<b>0.93</b>	<b>0.90</b>	<b>0.93</b>	<b>0.97</b>	<b>0.99</b>	<b>1.01</b>	<b>0.95</b>
MASE	Critical products	Holt	1.12	1.15	1.21	1.23	1.27	1.38	1.23
		Damped	<b>0.97</b>	<b>0.98</b>	<b>1.00</b>	<b>1.03</b>	<b>0.98</b>	<b>1.03</b>	<b>1.00</b>
		SES	1.02	1.04	1.07	1.06	1.10	1.08	1.06
MSSE	All products	Holt	1.13	1.14	1.18	1.23	1.32	1.46	1.24
		Damped	1.05	1.01	1.06	1.13	1.17	1.26	1.11
		SES	<b>0.93</b>	<b>0.88</b>	<b>0.90</b>	<b>1.03</b>	<b>1.06</b>	<b>1.10</b>	<b>0.98</b>
MSSE	Critical products	Holt	1.01	1.14	1.21	1.18	1.26	1.40	1.20
		Damped	<b>0.79</b>	<b>0.86</b>	<b>0.88</b>	<b>0.88</b>	<b>0.87</b>	<b>0.86</b>	<b>0.86</b>
		SES	0.92	0.96	0.90	0.98	1.08	0.95	0.96

Table 1. Average scaled error measures for the chemicals demand series. In each series, errors at all horizons were scaled by the one-step naïve error.

## Conclusions

Most previous research related to forecasting for supply chains has avoided the question of how forecasting methods should be chosen, and has been concerned instead with examining the negative effects of demand uncertainty. As discussed in Acar et al. (2009b), deterministic and often simplistic mathematical models are widely employed and are followed by multiple sensitivity analyses to judge the impact of demand uncertainty. The supply chain model in this paper is driven by actual daily demand data and integrates exponential smoothing, optimization, and simulation. We show that the choice of forecasting method makes a significant difference in operational performance.

Hyndman and Koehler's scaled error measures appear to be the best available options for measuring average forecast accuracy, but there is no relationship between operational performance and average accuracy across all products in this supply chain. In comparisons of average accuracy for inventory demands, the errors should logically be weighted by costs or customer service impact, an idea discussed in Syntetos et al. (2009). It may be possible to do this in a pure distribution inventory, where there is little interaction between inventory items, but it appears to be impossible to do so when modeling production, transportation, and distribution in the supply chain context.

Finally, one obvious question about this research is whether the damped trend is superior to the company's existing forecasting method. Anecdotal evidence suggests that the damped trend is an improvement, but the company has relied on purely subjective forecasts for many years and has not kept dependable records of forecasting performance.

## **Acknowledgment**

We wish to acknowledge Aris Syntetos, who made valuable comments on an earlier draft of this paper.

## **References**

Acar, Y., Kadipasaoglu, S., and Schipperijn, P. (2009a) A decision support framework for global supply chain modeling: an assessment of the impact of demand, supply, and lead-time uncertainties on performance, *International Journal of Production Research*, forthcoming.

Acar, Y., Kadipasaoglu, S., and Day, J. (2009b) Incorporating uncertainty in optimal decision making: Integrating mixed integer programming and simulation to solve combinatorial problems, *Computers and Industrial Engineering*, 56, 106-112.

Chandra, C., & Grabis, J. (2005). Application of multiple-steps forecasting for restraining the bullwhip effect and improving inventory performance under autoregressive demand, *European Journal of Operational Research*, 166, 337-350.

Dejonckheere, J., Disney, S.M., Lambrecht, M.R., & Towill, D.R. (2003). Measuring and avoiding the bullwhip effect: A control theoretic approach, *European Journal of Operational Research*, 147, 567-590.

Dejonckheere, J., Disney, S.M., Lambrecht, M.R., & Towill, D.R. (2004). The impact of information enrichment on the Bullwhip effect in supply chains: A control engineering perspective, *European Journal of Operational Research*, 153, 727-750.

Gardner, E.S., Jr., & McKenzie, E. (1985). Forecasting trends in time series, *Management Science*, 31, 1237-1246.

Gardner, E.S., Jr. (1990). Evaluating forecast performance in an inventory control system, *Management Science*, 36, 490-499.

Gardner, E. S. Jr., (2006) Exponential smoothing: The state of the art – Part II, *International Journal of Forecasting*, 22, 637-666

Holt, C.C. (2004). Forecasting seasonals and trends by exponentially weighted moving averages, *International Journal of Forecasting*, 20, 5-10.

Hyndman, R.J. & Koehler, A.B. (2006) Another look at measures of forecast accuracy, *International Journal of Forecasting*, 22, 679-688.

Law, A.N. & Kelton, W.D. (2000) *Simulation modeling and analysis* (3rd edition), New York: McGraw-Hill.

Sanders, N.R. & Graman, G.A. (2009) Quantifying costs of forecast errors: A case study of the warehouse environment, *Omega: The International Journal of Management Science*, 37, 116-125.

Snyder, R.D., Koehler, A.B., & Ord, J.K. (2002) Forecasting for inventory control with exponential smoothing, *International Journal of Forecasting*, 18, 5-18.

Syntetos, A.A., Nikoloupoulos, K., & Boylan, J.E. (2009) Judging the judges through accuracy-implication metrics: The case of inventory forecasting, *International Journal of Forecasting*, forthcoming.

Tiacci, L. & Saetta, S. (2009) An approach to evaluate the impact of interaction between demand forecasting method and stock control policy on the inventory system performances, *International Journal of Production Economics*, 118, 63-71.

Zhao, X., Xie, J., & Leung, J. (2002). The impact of forecasting model selection on the value of information sharing in a supply chain, *European Journal of Operational Research*, 142, 321-344.

Zhang, X. (2004). The impact of forecasting methods on the bullwhip effect, *International Journal of Production Economics*, 88, 15-27.