
ELSEVIER
VOLUME 28, ISSUE 4
OCTOBER–DECEMBER 2012
ISSN 0169-2070
international journal of forecasting
Special Section: Election Forecasting in Neglected Democracies
Guest Editors: Michael S. Lewis-Beck and Éric Bélanger
IIF
International Institute of Forecasters

International Journal of Forecasting 28 (2012) 842–848

Contents lists available at SciVerse ScienceDirect

International Journal of Forecasting

journal homepage: www.elsevier.com/locate/ijforecast

# Forecasting method selection in a global supply chain

Yavuz Acar [a], Everette S. Gardner Jr. [b,*]

[a] Bogazici University, Department of Management, Istanbul, Turkey

[b] University of Houston, Bauer College of Business, 334 Melcher Hall, Houston, TX 77204-6021, United States

ARTICLE INFO

*Keywords:*
Demand forecasting
Exponential smoothing
Inventory forecasting
Model selection
Simulation
Supply chain

ABSTRACT

In supply chains, forecasting is an important determinant of operational performance, although there have been few studies that have selected 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 will minimize the total supply chain cost. Tradeoff curves between total costs and customer service are used to compare exponential smoothing methods. The damped trend method produces the best tradeoffs.


## 1. Introduction

A comprehensive review of research in forecasting for supply chains is given by Fildes and Kingsman (2010), who conclude that there are few findings of any managerial importance. We agree. To ensure mathematical tractability, most researchers have assumed greatly simplified operating systems and cost structures. Furthermore, most have also failed to match the generation process for demand with the choice of a forecasting method. Thus, forecast errors have been compounded with misspecification errors, making it difficult to understand the effects of forecasting on efficiency, costs, inventory investment, or customer service levels. In a careful MRP simulation, Fildes and Kingsman set out to correct many of the fallacies in previous research. They found that the benefits of improved forecasting are considerably greater than the effects of choosing inventory decision rules, and that a misspecification of the forecasting method leads to increases in costs.

Fildes and Kingsman call for more empirical modelling of the supply chain that is grounded in observed practice, and that is the theme of this paper. We model the relationship between forecasting and operational performance in the supply chain of a global manufacturer of lubricants and fuel additives, products which are 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 for shipments between plants and to customers. Optimization depends on demand forecasts, supplied by exponential smoothing, and is done with a mixed integer program in order 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 & McKenzie, 1985). SES and the damped trend are obvious choices, given their long record of success in empirical studies (Gardner, 2006); the data suggested that the Holt method would not perform 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

* Corresponding author. Tel.: +1 713 743 4744; fax: +1 713 743 4940.
*E-mail address:* egardner@uh.edu (E.S. Gardner Jr.).

0169-2070/$ – see front matter 
doi:10.1016/j.ijforecast.2011.11.003

SES and Holt. It is interesting to contrast these results with traditional method selection based on average forecast accuracy measures; surprisingly, SES gave the best overall average accuracy.

## 2. 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, located in North and South America, Europe, and Asia. The 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 the 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 $\times$ 10 components).

The supply chain model, depicted in Fig. 1, 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 capacities, 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 the total variable supply chain cost, including 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, Kadipasaoglu, and Day (2009); Acar, Kadipasaoglu, and Schipperijn (2010). The models of Acar et al. were 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 using real data.

The simulation model at the second level executes the manufacturing plans on a daily basis, using the 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 was about 250,000 tons. As customer orders arrive, the demand is met by blending end products from the 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 the 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 that suggested in the first-level model, no further shipments can be scheduled until that model is run again.

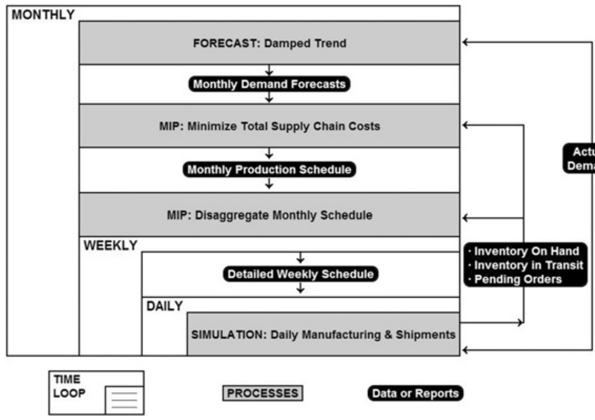


**Fig. 1.** The supply chain model.

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 of the 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 available to enable us to develop distributions of machine breakdowns, so we chose the following simulation procedure in consultation with maintenance and supply chain managers. The occurrence of breakdowns 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 a 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 by management to be the most important, is defined as the number of days an item is backordered times the backorder quantity.

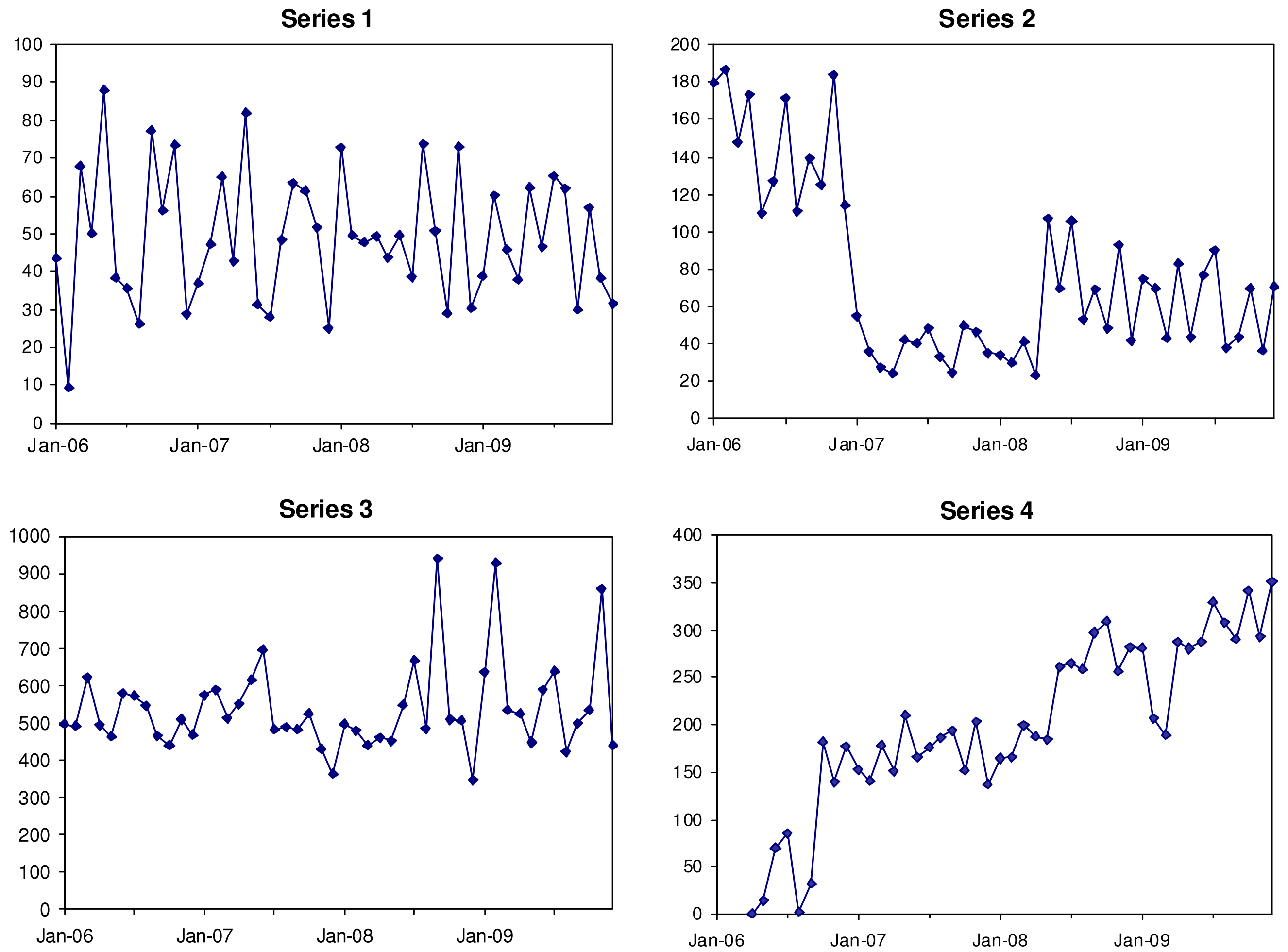


**Fig. 2.** Examples of monthly time series of component demands.

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 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, Koehler, and Ord (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—a common assumption but one that cannot be supported in most inventory systems.

The combined optimization and simulation model was run 300 times: 20 replications of transportation lead-times and machine breakdowns times three forecasting methods times five different 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.

## 3. 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 some 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 ill-behaved. Some series display a relatively constant level with extreme variance, as in Fig. 2, 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 for evaluating cost and service 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.

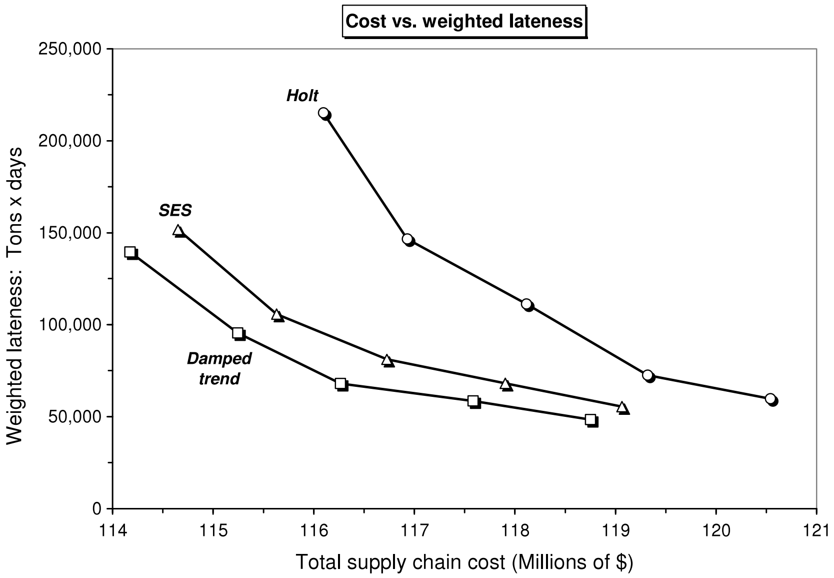


**Fig. 3.** Tradeoffs between total supply chain costs and weighted lateness (tons backordered times days on backorder) during the last year of operations.

However, it was not clear that the 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 during the fit periods.

Different procedures for initialization and fitting were tested, although they made no significant difference in cost or 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 the method components (level and trend), when the methods were fitted so as to minimize the mean absolute error, and when the methods were initialized using only the first six months or the first year of data.

## 4. Tradeoff analysis

Gardner (1990) recommended the use of tradeoff curves for evaluating the operational performance of forecasting methods, and we followed that example here. For the last year of operations, Fig. 3 gives a tradeoff curve for total supply chain costs vs. average weighted lateness, and Fig. 4 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, about 41,000 customer orders were processed during the last year of operations, and the percentages of backorders were rather large, ranging from about 7%–19% for the damped trend, for example. We were concerned about 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 Figs. 3 and 4, 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 a 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 Figs. 3 and 4, as costs and safety stocks increase, the differences between methods become smaller, as should be expected, although the differences are always significant.

Production costs account for an average of 90% of the 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 operate 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 are sensitive to the choice of forecasting method. The consequences of forecast errors are extremely complicated, and there is no simple explanation as to 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 for others as well.

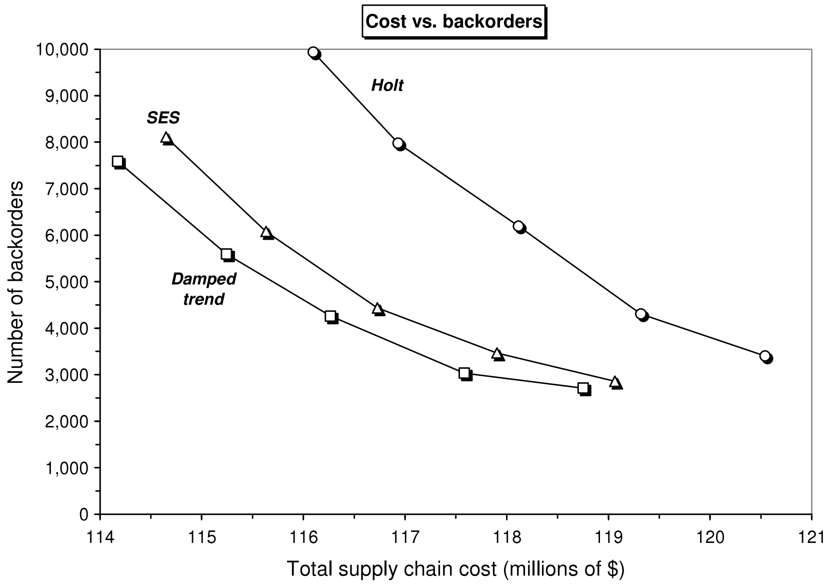


**Fig. 4.** Tradeoffs between total supply chain costs and numbers of backorders during the last year of operations.

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 stock shortages at a later date. When backorders are filled, emergency shipments are necessary, with transportation costs which are 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 serious backorder problems. 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 which interact with each other in allocating production capacity, and the system is dynamic with monthly updates. This means that forecast errors are frequently 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 is robust, a property demonstrated in many previous studies of forecast accuracy. The damped trend also performed especially well for a group of four critical components that account for about 3/4 of total production.

## 5. Average forecast accuracy

It is interesting to compare the average forecast accuracy of the exponential smoothing methods with their customer service and cost performance. For horizons 1–6, Table 1 gives the mean absolute percentage error (MAPE), mean percentage error (MPE), mean absolute scaled error (MASE), and mean square scaled error (MSSE). The MAPE and MPE are seriously misleading in these series because they vary drastically in scale, with some observations near zero. We therefore 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.

Judged by the MAPE, MASE, and MSSE measures, SES is the most accurate method for all products and forecast horizons, followed by the damped trend and Holt. For the group of four critical products mentioned above, the damped trend is the most accurate method, which supports the service and cost tradeoffs above, although we can see no way to predict service or cost from the average accuracy.

Judged by the MPE values over all products, all methods had a tendency to over-forecast, and the Holt method was the least biased while SES was the worst. For the critical products, the rankings are reversed, with Holt as the worst and SES as the best. Sanders and Graman (2009) argue that bias in forecasts is more important than the average accuracy in determining costs, and that there is some optimal amount of bias for a given cost structure. Fildes and Kingsman (2010) are sharply critical of this line of research, and it is not clear how Sanders and Graman's arguments are relevant to our study. The MPE results in Table 1 are difficult to explain except to say that they are distorted by enormous percentage outliers in series with observations near zero. It is easy to change the averages and ranking of the methods by removing a few selected series. In summary, the average forecast accuracy results were unhelpful, and of no interest to management.

## 6. Conclusions

There appears to be no previous research on forecasting method selection based on operational performance in a real supply chain. 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

**Table 1**
Average forecast error measures for the chemicals demand series. For the MASE and MSSE measures, errors at all horizons were scaled by the one-step-ahead naïve error.

| | | | Horizon | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | | 1 | 2 | 3 | 4 | 5 | 6 | All |
| MAPE | All products | Holt | 38.7 | 39.4 | 41.3 | 43.6 | 45.4 | 49.1 | 42.9 |
| | | Damped | 35.6 | 35.6 | 37.6 | 39.3 | 40.7 | 42.8 | 38.6 |
| | | SES | **30.5** | **30.6** | **31.8** | **33.6** | **35.0** | **35.3** | **32.8** |
| MAPE | Critical products | Holt | 16.0 | 16.8 | 17.3 | 17.0 | 17.8 | 18.9 | 17.3 |
| | | Damped | **13.0** | **13.6** | **13.8** | **13.9** | **13.4** | **14.0** | **13.6** |
| | | SES | 13.3 | 13.9 | 14.3 | 14.2 | 14.3 | 14.2 | 14.0 |
| MPE | All products | Holt | **−3.1** | **−3.4** | **−3.4** | **−2.7** | **−3.4** | **−2.7** | **−3.1** |
| | | Damped | −4.3 | −4.9 | −5.4 | −5.4 | −6.6 | −6.5 | −5.5 |
| | | SES | −11.7 | −13.2 | −13.7 | −14.5 | −15.7 | −15.6 | −14.1 |
| MPE | Critical products | Holt | −8.4 | −10.4 | −10.5 | −9.5 | −11.2 | −11.2 | −10.2 |
| | | Damped | −2.6 | −4.3 | −4.0 | −2.6 | −3.5 | **−2.5** | −3.2 |
| | | SES | **0.0** | **−1.1** | **0.0** | **2.1** | **2.3** | 4.3 | **1.3** |
| 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 | **0.93** | **0.90** | **0.93** | **0.97** | **0.99** | **1.01** | **0.95** |
| MASE | Critical products | Holt | 1.12 | 1.15 | 1.21 | 1.23 | 1.27 | 1.38 | 1.23 |
| | | Damped | **0.97** | **0.98** | **1.00** | **1.03** | **0.98** | **1.03** | **1.00** |
| | | 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 | **0.93** | **0.88** | **0.90** | **1.03** | **1.06** | **1.10** | **0.98** |
| MSSE | Critical products | Holt | 1.01 | 1.14 | 1.21 | 1.18 | 1.26 | 1.40 | 1.20 |
| | | Damped | **0.79** | **0.86** | **0.88** | **0.88** | **0.87** | **0.86** | **0.86** |
| | | SES | 0.92 | 0.96 | 0.90 | 0.98 | 1.08 | 0.95 | 0.96 |

significant difference to both the customer service and cost tradeoffs available to management.

Hyndman and Koehler's scaled error measures are the best available options for measuring the average forecast accuracy, but there is no relationship between operational performance and average accuracy across all products in this supply chain. Syntetos, Nikoloupoulos, and Boylan (2010) argue that, in comparisons of average accuracy for inventory demands, the errors should be weighted by the cost or customer service impact. We agree that this should be done in a pure distribution inventory, where there are few interactions between inventory items, but it is difficult to do so when modeling production, transportation, and distribution in the supply chain context. The consequences of forecast errors are complex because there are powerful interactions between products competing for the same production capacity. These interactions lead us to the conclusion that forecasting must be evaluated at the aggregate level in the form of cost-service tradeoff curves for the entire supply chain.

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 no clearly defined forecasting method at present, so there is no real basis for comparison. The company has relied on purely subjective forecasts for many years, and there are no reliable records of forecast values, when forecasts were made, or how they were made. Our experience is that this is not unusual.

## Acknowledgments

We wish to acknowledge Aris Syntetos, who made valuable comments on an earlier draft of this paper and helped us clarify the description of the supply chain model.

## References


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

Acar, Y., Kadipasaoglu, S., & Schipperijn, P. (2010). 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*, *48*, 3245–3268.

Fildes, R., & Kingsman, B. G. (2010). Incorporating demand uncertainty and forecast error in supply chain planning models. *Journal of the Operational Research Society*, *62*, 483–500.

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.

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

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 ed.). 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. (2010). Judging the judges through accuracy-implication metrics: the case of inventory forecasting. *International Journal of Forecasting*, *26*, 134–143.

**Yavuz Acar** is Assistant Professor in the Department of Management at Bogazici University in Istanbul, where he teaches supply chain management. His research interests include mathematical programming, supply chain modeling, and demand forecasting.

**Everette S. Gardner, Jr.**, is Professor and Director of the Honors Program in the Bauer College of Business at the University of Houston, where he teaches Honors colloquia in forecasting and supply chain management. His research interests include exponential smoothing, supply chain modeling, and forecast evaluation.