The Effect of Planning Horizon on the Effectiveness of What-If Analysis

JE KOTTEMANN

University of Michigan, USA

WE REMUS

University of Hawaii, USA

Although what-if analysis is among the most popular decision support methods, empirical evidence indicates that it does not predictably improve decision making. This suggests that the effectiveness of what-if analysis is, in part, contingent on characteristics of what-if models. However, extremely little research has been conducted to identify and assess these characteristics. A between-subjects experiment was conducted to determine the effect of one potentially critical characteristic of what-if tools used for planning: the planning horizon reflected in the model. The authors hypothesized that the "temporal framing" embodied by what-if models with different planning horizons would result in differential effects on short-run and long-run performance. The hypothesis was supported: subjects provided a multi-period what-if model performed significantly better in the long-run than those provided a one-period model. In addition, both what-if groups out-performed the control group, but only in the short-run. Surprisingly, the control group out-performed both what-if groups in the long-run. These results suggest that the effectiveness of what-if analysis is contingent on subtle, yet fundamental, characteristics of what-if models. Future research that identifies other key characteristics will help provide modelers with specific, practical guidelines for effective model construction.

Key words—decision support systems, production planning

INTRODUCTION

WHAT-IF ANALYSIS, among the most popular decision support methods, allows decision makers to manipulate parameters of decision models in an attempt to make better decisions. For example, investment bankers use what-if analysis to "manipulate" factors such as future interest rates and porfolio diversifications in order to formulate investment strategies. Despite the popularity of what-if analysis, research comparing what-if analysis to alternative decision support methods and to unaided decision making has shown that what-if analysis sometimes generates positive performance effects (e.g. [1, 13]), sometimes no effects (e.g. [3, 4]) and sometimes negative effects (e.g. [8, 9]).

The equivocality of these findings suggests that the effectiveness of what-if analysis is contingent on characteristics of decision makers, decision tasks, and what-if modeling and analysis itself. Research into the latter is needed to discover the specific characteristics of what-if models that affect decision making performance. However, such research is almost nonexistent. One exception is the study by Dos Santos and Barriff [2] who found that what-if analysis was most effective when the outcome variables in the model were given as deviations from budget rather than as absolute values. In the experiment reported here, we investigate the effects of differences in one potentially critical characteristic of what-if models used for planning: the planning horizon reflected in the what-if model.

Decision makers in practice are often criticized for adopting short-run strategies, wherein they optimize in the short-run at the expense of long-run performance. Some authors (e.g. [5, 10]) have offered anecdotal evidence suggesting that part of this problem stems from the influence of potentially biasing analytic tools. Here, we investigate whether the planning horizon reflected in a what-if model affects decision making performance over time. More specifically, we investigate whether the planning horizon reflected in a what-if analysis tool significantly influences decision makers' when they formulate and execute decision strategies, resulting in differential effects on short-run and long-run performance.

METHOD

The experimental task

The experimental task is based on the Holt, Modigliani, Muth, and Simon (HMMS) formulation of the production planning problem [6, 7]. In this formulation, a decision maker faces uncertain demand and must choose a production level and a workforce level over a series of periods. The objective is to minimize cumulative total costs. The cost function for a given new period is the sum of three quadratic cost components:

Workforce level change cost =
    64.3 ((Current workforce
      − New workforce)^2)

Worker overtime/idletime cost =
    0.8 (((New workforce × 5.67)
      − New production)^2)

Cost for nonoptimal inventory =
    0.02 ((Current inventory
      + New production
      − New demand − 320)^2)

where 5.67 is the number of units each worker can produce each period, and 320 is the optimal safety stock level. Notice that the cost components are highly interrelated and that one cost component can be avoided only at the expense of another. For example, if the workforce is set artificially low relative to production, then higher worker overtime cost is incurred. If, in turn, production is set artificially low relative

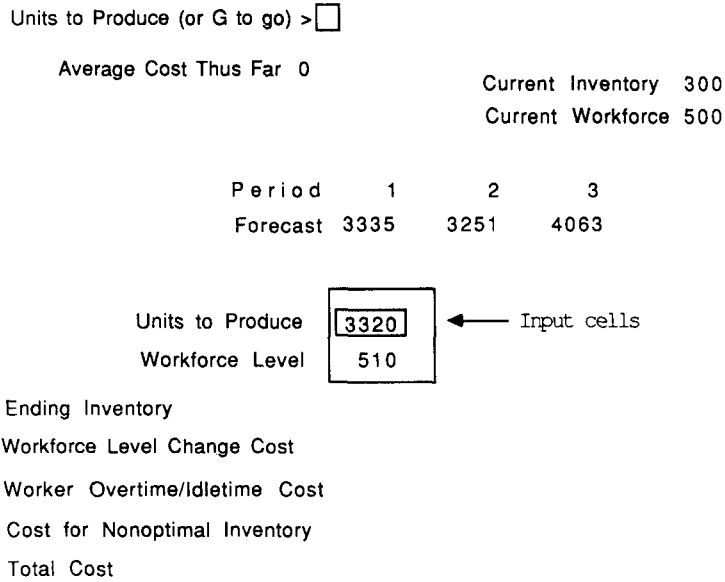to demand, then inventory outage cost is driven higher.

The demand trend used in the experiment increased linearly. Demand began at 2500 units and increased at a rate of 100 units per period. This "unadjusted" demand was randomized ± 200 units to generate the *actual demand*. Subjects' social security numbers were used to seed the random number generator.

After four practice periods, subjects made production and workforce decisions for each of 24 periods. All subjects received the following information for each period: the current workforce level, the current inventory level, and demand forecasts for each of the next three periods. Forecasts were generated by further randomizing the actual demands. The choice of three periods for forecasting was based on Moskowitz and Miller's [11] research demonstrating its superiority over shorter forecast horizons in this production planning task; the forecast error level was set to correspond to their intermediate level.

In order to assess the effects of short-run and long-run strategies on performance, we conducted a set of comparative simulations. The first simulations involved a strategy that, for each decision period, minimized cost in that particular period (termed the short-run strategy). The second involved an exponential smoothing heuristic that attempted to track future demand using the three periods of forecasts, to produce at a level to meet demand over time, and to hire the proper number of workers given the production level. These simulations revealed that, while the short-run strategy incurred lower costs in the initial decision periods, it incurred significantly higher costs in the later periods. We expect, then, that if a what-if model supports analysis on only the current, upcoming period, it will result in higher long-run costs relative to a model that supports multi-period analysis. Details of the what-if models used in the experiment and our hypotheses are given below.

The decision aids

The experiment was a between-group design with three treatments: a control group, a one-period what-if model, and a three-period what-if model. The task simulator discussed above was augmented with a full-screen user interface with three levels of functionality corresponding to the

Units to Produce (or G to go) >☐

Average Cost Thus Far 0

Current Inventory 300
Current Workforce 500

Period   1   2   3
Forecast   3335   3251   4063

Units to Produce   │3320│   ◄——— Input cells
Workforce Level   510

Ending Inventory

Workforce Level Change Cost

Worker Overtime/Idletime Cost
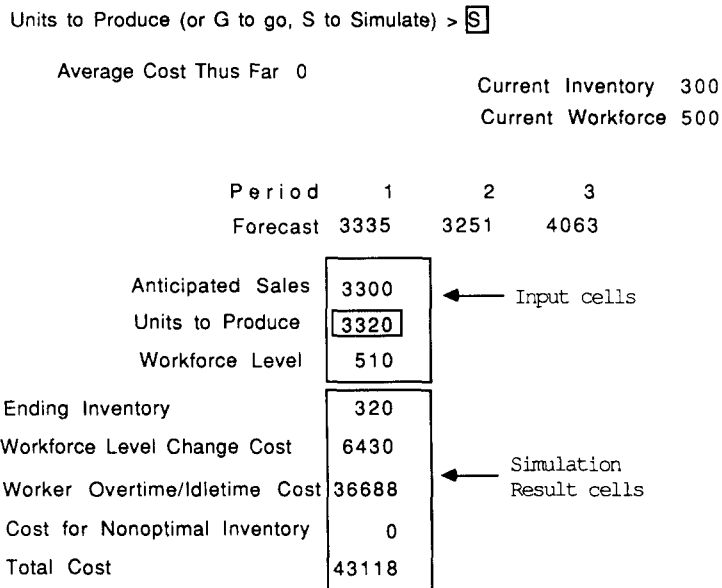
Cost for Nonoptimal Inventory

Total Cost

Worker Productivity is always 5.67 Units Per Worker Per Period
The Optimal Level of Inventory is always 320 Units

Fig. 1. The control group's interface.

three treatments. All interfaces were pre-tested using colleagues, graduate and undergraduate students to assess and improve ease of use.

Figure 1 shows the control group's screen layout. The lower lines on the screen displayed the worker productivity index and the optimal level at which to maintain inventory. Both these factors remained constant throughout the task.

Units to Produce (or G to go, S to Simulate) > S

Average Cost Thus Far 0

Current Inventory 300
Current Workforce 500

Period   1   2   3
Forecast   3335   3251   4063

Anticipated Sales   3300   ◄——— Input cells
Units to Produce   │3320│
Workforce Level   510

Ending Inventory   320
Workforce Level Change Cost   6430   Simulation
Worker Overtime/Idletime Cost 36688   ◄——— Result cells
Cost for Nonoptimal Inventory   0
Total Cost   43118

Worker Productivity is always 5.67 Units Per Worker Per Period
The Optimal Level of Inventory is always 320 Units

Fig. 2. The one-period what-if interface.

Units to Produce (or G to go, S to Simulate) > $\boxed{\text{S}}$

Average Cost Thus Far 0

Current Inventory 300
Current Workforce 500

P e r i o d	1	2	3
Forecast	3335	3251	4063
Anticipated Sales	3300	3250	4000	◄—— Input cells
Units to Produce	$\boxed{3320}$	3300	4100	
Workforce Level	510	515	525	
Ending Inventory	320	370	470	
Workforce Level Change Cost	6430	1608	6430	
Worker Overtime/Idletime Cost	36688	28872	252338	◄—— Simulation
Cost for Nonoptimal Inventory	0	206	1856	Result cells
Total Cost	43118	30686	260624	

Worker Productivity is always 5.67 Units Per Worker Per Period
The Optimal Level of Inventory is always 320 Units

Fig. 3. The three-period what-if interface.

The upper right corner displayed the current workforce and inventory levels. The upper left displayed the average total cost thus far incurred. The upper middle of the screen displayed the forecasts generated for the next three upcoming periods. The middle of the screen was used to accept subjects' production and workforce level inputs.

ENTER RETURN TO PROCEED TO THE NEXT DECISION PERIOD

Average Cost Thus Far 85846

Actual Sales 3575

Current Inventory 245
Current Workforce 510

P e r i o d	1	2	3
Forecast	3543	3753	3722

Units Produced   3520
Workforce Level   510

Workforce Level Change Cost   6430
Worker Overtime/Idletime Cost 78952
Cost for Nonoptimal Inventory   464
Total Cost   85846

Worker Productivity is always 5.67 Units Per Worker Per Period
The Optimal Level of Inventory is always 320 Units

Fig. 4. Actual results screen (all treatment groups).

Subjects in the one-period what-if group received, in addition to the above, a what-if function which was embedded into the task simulator (see Fig. 2). With this function, subjects could simulate any number of scenarios per period to assess cost ramifications. They could enter levels for anticipated sales demand, workforce and production levels, and have the system calculate the resulting costs. The formulas used in the what-if function were the actual cost formulas underlying the task.

As shown in Fig. 3, subjects in the three-period what-if group were able to perform what-if analysis on the next upcoming three decision periods. While their actual decisions applied only to the next upcoming period, the ability to "plan ahead" was expected to help their long-run performance relative to the one-period what-if group. This result was expected even though the one-period what-if group also received three period forecasts.

In the short-run, we expect both what-if groups to out-perform the control group for two reasons. First, use of what-if analysis allows subjects to experiment freely with the task, and so should help them learn the task more rapidly. This conjecture is supported by Sharda et al.'s [13] study, in which the what-if treatment group gained its primary performance advantage over the control group in the first few decision periods. Second, if what-if analysis does encourage short-run strategies (notably for the one-period what-if group), then a performance advantage should be expected in the initial decision periods. Hypotheses regarding the long-run performance of the control group relative to the what-if groups are less clear-cut. If indeed the one-period group adopts a short-run strategy, then it might be expected that the control group will out-perform them in the long-run. It is less clear whether the three-period what-if group's performance will differ from the control's.

Subjects entered the "go" command whenever they were ready to process the decision they had entered. The system then performed checks for gross numeric range errors (e.g. producing 20,000 units with a workforce of 0) and displayed error messages as appropriate. In any event, subjects were asked to confirm the "go" command before the decision was processed. Upon confirmation, the results were computed and displayed. As shown in Fig. 4, the display included the actual demand for the period, the new inventory and workforce levels, workforce change cost, worker over/idle time cost, inventory over/outage cost, total cost for the period, and the updated average total cost thus far incurred.

Subjects and procedure

There were 39 subjects from an MBA course in operations research and information systems; they received nominal class credit for their participation. They were randomly assigned to treatments. All subjects had end-user computing experience including the use of a spread-sheet package. A week prior to the experiment, one of the authors lectured the entire group of subjects on production processes. In this lecture, the general nature of the production planning task was overviewed in the spirit of the experimental task description given above. The specific values of the cost model coefficients were not given. They were told that the objective was to minimize cost in the long-run.

The subjects were scheduled to arrive at the computing lab every 10 minutes. They were directed to a terminal, and one of the authors reintroduced them individually to the production planning task and instructed them in the use of the specific software to be used. These orientation sessions were based upon pre-tested scripts. The what-if groups were given a demonstration of the what-if function.

Subjects were given four practice periods and then performed the 24 "real" periods. Subjects were told that they could take as long as they wanted to complete the task; they took from 40 to 180 minutes. They were not allowed to interact. Also, given the seating arrangement and staggered arrival times, it was difficult for subjects to compare how long they were spending on the task relative to others. The subjects appeared highly motivated (e.g. they were eager to find out how they did).

Results

In order to investigate performance differences between groups as well as over time, we partitioned cost performance for each group into four phases. The periods in each phase are 2–6, 7–12, 13–18, and 19–24. Period 1 is omitted from the analysis because all subjects had the same starting conditions. Table 1 presents the results for each of the four phases using rank transformed analysis of variance.

Table 1. Costs differences over time using rank transformed analysis of variance (means with the same Duncan grouping letter are not significantly different)

Duncan grouping	Mean cost rank	Treatment
Phase 1 (Periods 2–6)		
A	131.94	Control
B	106.51	One-period what-if
B	103.21	Three-period what-if
P = 0.01		
Phase 2 (Periods 7–12)		
A	118.01	Control
A	115.38	One-period what-if
A	110.18	Three-period what-if
P = 0.75		
Phase 3 (Periods 13–18)		
A	141.17	One-period what-if
B	112.58	Three-period what-if
C	91.81	Control
P = 0.0000		
Phase 4 (Periods 19–24)		
A	144.08	One-period what-if
B	115.83	Three-period what-if
C	85.86	Control
P = 0.0000		

During phase 1, no significant difference exists between the one-period and three-period what-if groups. However, both what-if groups incurred significantly lower costs than the control group ($P = 0.01$). Thus, there is evidence that what-if analysis improved short-run task performance. However, this difference has disappeared in phase 2 ($P = 0.75$).

In phases 3 and 4, there are significant performance differences between all three groups. As predicted, the three-period what-if group incurred significantly lower costs than the one-period group in the later decision periods ($P = 0.0000$). Surprisingly, however, the control group significantly out-performed both what-if groups in the final two phases ($P = 0.0000$).

In order to test for differences in long-run cost trends, we used the general linear model test [12, pp. 328–367]. In this test four regressions were simultaneously fitted for periods 13–24 (one for each treatment and one for the combined data); cost was estimated as a function of the period number; the slope of these lines is the cumulative cost change per period. Based on this test, the cost trends in periods 13–24 for the control, one-period, and three-period what-if groups are $202, $27,077, and $13,481 per period, respectively. They were significantly different ($P < 0.01$); *post hoc* analysis showed each to be significantly different from the others ($P < 0.01$). Taken together, the above statistical tests indicate that the long-run performance differences are significant both when viewed as cost amounts and when viewed as cost trends.

DISCUSSION AND CONCLUSION

Together with the findings of Dos Santos and Barriff [2], discussed in the introduction, the present study suggests that the effectiveness of what-if tools is contingent on relatively subtle, yet fundamental, characteristics of how the tools are structured. Dos Santos and Barriff [2] found that what-if analysis was most effective when the outcomes in the model were given as deviations from budget rather than as absolute values. In short, their results suggest that the "outcome framing" reflected in a what-if analysis tool affects performance. The present study shows that the planning horizon of a what-if model, and the corresponding "temporal framing" that the planning horizon imparts, can significantly affect decision making performance.

As hypothesized, the three-period what-if model led to better performance than the one-period model in the later decision phases. Also as expected, both what-if groups out-performed the control group in the first decision phase. However, the control group out-performed both what-if groups in the later decision periods. While we did speculate that the control group might out-perform the one-period what-if group in the long-run, we did not hypothesize that this would also be the case for the three-period what-if group. It is possible that subjects provided a three-period what-if analysis tool still tended to use relatively short-run strategies, and that providing subjects with a what-if tool that incorporates still more periods would reverse this effect. However, it is unclear just how many periods would be required for the long-run performance of what-if users to exceed unaided subjects. It is clear that what-if analysis itself, and the planning horizons involved, can have significant, differential effects on short-run and long-run performance.

The equivocality of findings in research comparing what-if analysis to alternative decision aiding methods and to unaided decision making (e.g. [1, 3, 8, 9, 13]) suggests that the effectiveness of what-if analysis is contingent on, among other things, characteristics of what-if modeling and analysis tools.

While there is evidence that "outcome framing" and "temporal framing" are two such characteristics, future research is needed to verify their effects. Future research is also needed to identify other key characteristics. The

popularity of what-if analysis underscores the importance of research designed to identify key performance characteristics of what-if analysis tools. Such research will help provide modelers with specific, practical guidelines for effective tool construction.

REFERENCES

1. Benbasat I and Dexter AS (1982) Individual differences in the use of decision support aids. *J. Accng Res.* **20**(1), 1–11.
2. Dos Santos BL and Bariff ML (1988) A study of user interface aids for model-oriented decision support systems. *Mgmt Sci.* **34**(4), 461–468.
3. Fripp J (1985) How effective are models? *Omega* **13**(1), 19–28.
4. Goslar M, Green G and Hughes T (1986) Decision support systems: An empirical assessment for decision making. *Decis. Sci.* **17**(1), 79–91.
5. Hayes RH and Abernathy WJ (1980) Managing our way to economic decline. *Harvard Bus. Rev.* (August), 67–77.
6. Holt CC, Modigliani F and Muth JF (1956) Derivation of a linear decision rule for production and employment. *Mgmt Sci* **2**(2), 159–177.
7. Holt CC, Modigliani F, Muth JF and Simon HA (1960) *Planning—Production, Inventories, and Work Force.* Prentice–Hall, Englewood Cliffs, NJ.
8. Kottemann JE and Remus WE (1987) Evidence and principles of functional and dysfunctional decision support systems. *Omega* **15**(2), 135–144.
9. Kottemann JE and Remus WE (1991) The effects of decision support systems on performance. In *Environments for Supporting Decision Processes* (Edited by Sol HG and Vecsenyi J), pp. 203–214. North-Holland, New York.
10. Lenz RT and Lyles MA (1985) Paralysis by analysis: Is your planning system becoming too rational? *Long Range Planning* **18**, 64–72.
11. Moskowitz H and Miller JG (1975) Information systems and decision systems for production planning. *Mgmt Sci* **22**(3), 359–371.
12. Neter J, Wasserman W and Kutner MH (1983) *Applied Linear Regression Models.* Irwin, Homewood, IL.
13. Sharda R, Barr SH and McDonnell JC (1988) Decision support system effectiveness: A review and an empirical test. *Mgmt Sci.* **34**(2), 139–159.

ADDRESS FOR CORRESPONDENCE: *JE Kottemann. School of Business Administration, University of Michigan, Ann Arbor, MI 48109-1234, USA.*