TECHNICAL NOTES

SENSITIVITY OF WEIGHTS IN COMBINING FORECASTS

ROBERT L. WINKLER

Duke University, Durham, North Carolina

ROBERT T. CLEMEN

University of Oregon, Eugene, Oregon

(Received November 1989; revision received December 1990; accepted March 1991)

In the combination of forecasts, weighted averages that attempt to take into account the accuracy of the forecasts and any dependence among forecasts tend to perform poorly in practice. An important factor influencing this performance is the sensitivity, or instability, of the estimated weights used to generate the combined forecast. The intent of this paper is to look at this instability via graphs and the sampling distribution of the weights. Results are developed for the combination of two forecasts and extended to the *m*-forecast case by viewing the *m*-forecast case as a sequence of two-forecast combinations.

A model used frequently to combine forecasts involves a weighted average of the individual forecasts that can be interpreted as a variance-minimizing forecast (Newbold and Granger 1974) or as a posterior mean in a Bayesian framework (Winkler 1981). The weights used to combine the forecasts are based on a covariance matrix, thus accounting for differences in the accuracy of the forecasts (through the variances) and dependence among the forecasts (through the correlations). Applications of this model in real-world forecasting situations have met with somewhat mixed results (e.g., see Clemen 1989). In practice, the variances and correlations are now known and must be estimated. When forecast errors are highly correlated, as is often the case, estimation of the covariance matrix can result in unstable combination weights. Bunn (1985) showed how the variance ratio and correlation between two forecasters' errors interact to undermine robustness. Clemen and Winkler (1986) provided examples of unstable weights from their analysis, while Kang (1986) illustrated instability through a simulation experiment.

The objective of this note is to investigate the sensitivity of weights in combining forecasts through the derivation of the sampling distribution of the combining weights and through various graphs. The combination of two forecasts is addressed in Section 1, and extensions to the case of more than two forecasts are presented in Section 2. In Section 3 we summarize our findings.

1. COMBINING TWO FORECASTS

1.1. Graphs for Combining Two Forecasts

Suppose that a decision maker wants to forecast θ and has obtained two forecasts, f_1 and f_2, of θ. If the forecast errors $f_1 - \theta$ and $f_2 - \theta$ have zero means, standard deviations σ_1 and σ_2, and correlation ρ, the combined forecast is

$$f_c = w_1 f_1 + (1 - w_1) f_2, \quad (1)$$

where $w_1 = (1 - \rho\phi)/(1 + \phi^2 - 2\rho\phi)$ and $\phi = \sigma_1/\sigma_2$. The variance of the error $f_c - \theta$ is $\sigma_c^2 = \sigma_1^2 (1 - \rho^2)/(1 + \phi^2 - 2\rho\phi)$.

The weight w_1 is graphed as a function of ρ for chosen values of ϕ in Figure 1. When the error variances are equal, $\phi = 1$ and $w_1 = 0.5$ for all ρ. With unequal error variances, however, w_1 shifts away from 0.5 as ρ increases. Furthermore, this movement away from 0.5 occurs at an increasing rate as ρ increases. These results can be verified by differentiating w_1 twice with respect to ρ. When $\phi > \rho > 0$ or $\phi^{-1} > \rho > 0$, one of the weights (w_1 or $1 - w_1$) is greater than one

Subject classification: Forecasting: combining forecasts.
Area of review: DECISION ANALYSIS, BARGAINING AND NEGOTIATION.

0030-364X/92/4003-0609 $01.25

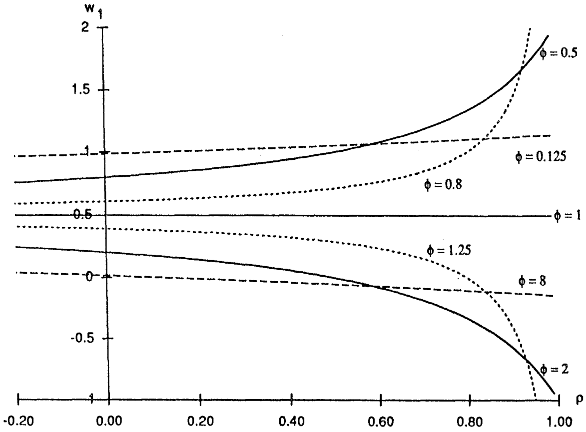

Figure 1. Weight w_1 assigned to the first forecast as a function of ρ, the correction between forecast errors, for selected values of $\phi = \sigma_1/\sigma_2$.

and the other is negative. Weight w_1 appears to be more sensitive to changes in ρ for high values of ρ.

In Figure 2, contours of constant weight w_1 are given in (ϕ, ρ) – space. Note, in particular, how these contours converge to a single point as $(\phi, \rho) \rightarrow (1, 1)$. It appears that w_1 is most sensitive to changes in ϕ in the region near (1, 1). Also, when ϕ is near one but not equal to one, w_1 is highly sensitive to changes in ρ. With $\phi = 1.1$, for example, it takes a much smaller change in ρ to move from $w_1 = 0.25$ to $w_1 = -0.25$ than it does with $\phi = 3$. Furthermore, the error variance σ_c^2 can also fluctuate widely when ρ is high and ϕ is near one.

1.2. Variability of Estimated Weights

The parameters σ_1, σ_2, and ρ are not known in practice, and they are typically estimated from past data. The resulting weight $\hat{w}_1$ is an estimate of w_1, and

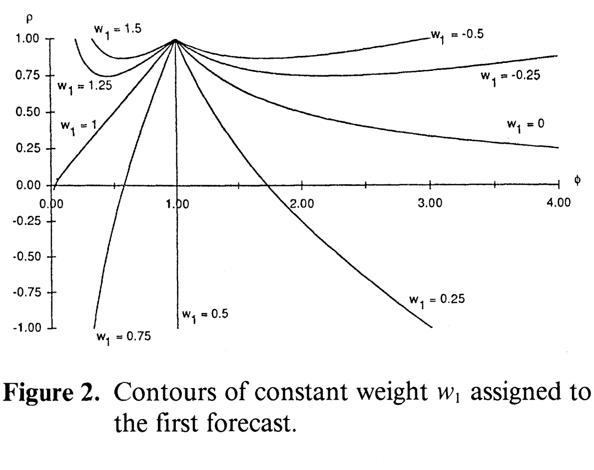

Figure 2. Contours of constant weight w_1 assigned to the first forecast.

sampling fluctuations will cause $\hat{w}_1$ to deviate from w_1. To study these fluctuations, we consider the sampling distributions of $\hat{w}_1$ given σ_1, σ_2, and ρ.

The model can be written in the form

$$\theta - f_2 = w_1(f_1 - f_2) + e, \tag{2}$$

where $e = \theta - f_c$ has mean zero and variance σ_c^2. When w_1 is estimated by replacing the parameters σ_1, σ_2, and ρ by their sample counterparts, the resulting estimator $\hat{w}_1$ is a least-squares (and maximum-likelihood, given a normality assumption) estimator of w_1 in a regression model represented by (2). The sampling distribution of $\hat{w}_1$ given σ_1, σ_2, and ρ is a Pearson Type VII distribution with density

$$g(\hat{w}_1 \mid \sigma_1, \sigma_2, \rho)$$
$$\propto \{[\sigma_2^{*2}(1 - \rho^{*2})/\sigma_1^{*2}] + (\hat{w}_1 - w_1)^2\}^{-n/2}, \tag{3}$$

where

$$\sigma_1^{*2} = V(f_1 - f_2) = \sigma_1^2 + \sigma_2^2 - 2\rho\sigma_1\sigma_2,$$

$$\sigma_2^{*2} = V(\theta - f_2) = \sigma_2^2,$$

$$\rho^* = \mathrm{Cov}(\theta - f_2, f_1 - f_2)/\sigma_1^*\sigma_2^* = (\sigma_2^2 - \rho\sigma_1\sigma_2)/\sigma_1^*\sigma_2^*,$$

and n is the sample size for the data used to estimate w_1 (see Kendall and Stuart 1963, p. 392).

The standard deviation

$$\sigma_{\hat{w}_1} = \sqrt{\phi^2(1 - \rho^2)/(n - 3)(1 + \phi^2 - 2\rho\phi)^2}$$

is graphed for different values of ϕ as a function of ρ in Figure 3. The two vertical axes show values of $\phi_{\hat{w}_1}$ for $n = 30$ and $n = 100$. Note that the variability of $\hat{w}_1$ is particularly great when ϕ is near one and ρ is high but not quite one. For example, if $\phi = 0.9$ and $\rho = 0.8$, then $w_1 = 0.76$ and $\sigma_{\hat{w}_1} = 0.28$ when $n = 30$, and $w_1 = 0.76$, and $\sigma_{\hat{w}_1} = 0.15$ when $n = 100$.

Relating the sampling distribution in (3) to the t distribution, we find that if $\phi = 0.9$ and $\rho = 0.8$, the probability is 0.95 that $\hat{w}_1$ will be in the interval (0.21, 1.31) when $n = 30$ and (0.46, 1.06) when $n = 100$. These are very wide ranges of possible weights, including (even with $n = 100$) weights less than 0.50 (more weight on the second forecast than the first) and higher than one (negative weight on the second forecast).

Even when $\phi = 1.0$, implying that the weights should be equal, there can be a good chance that one of the estimated weights is negative if the correlation is high. With $n = 30$, for example, the probability of a negative weight when $\phi = 1.0$ is 0.40 when $\rho = 0.95$, 0.23 when $\rho = 0.90$, and 0.09 when $\rho = 0.80$; this probability drops to 0.006 when $\rho = 0.50$. Shifting ϕ away from one increases the chance of a negative weight because w_1 moves away from 0.50. When $n = 30$ and

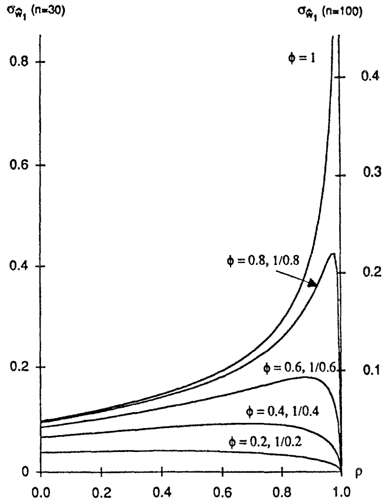

Figure 3. Standard deviation of the estimated weight $\hat{w}_1$ as a function of ρ, the correlation coefficient between forecast errors, for selected values of $\phi = \sigma_1/\sigma_2$. Note that the curves for $\phi = \sigma_1/\sigma_2$ are the same as those for $\phi = \sigma_2/\sigma_1$ because of the symmetry of the problem of combining two forecasts.

$\phi = 0.80$, the probability of one of the two weights being negative is 0.88 when $\rho = 0.90$, 0.50 when $\rho = 0.80$, and 0.23 when $\rho = 0.70$.

As these numerical examples and the curves in Figure 3 demonstrate, $\hat{w}_1$ can be highly variable, particularly when ϕ is near one and ρ is high but not quite one. But this is exactly the region where σ_c^2 is particularly sensitive, and the estimate of the accuracy of the combined forecast can be expected to be quite variable. It is also exactly the region consistent with empirical estimates of ϕ and ρ (e.g., Clemen and Winkler 1986).

2. COMBINING MORE THAN TWO FORECASTS

2.1. Combining Sequentially

Once again, the decision maker wants to forecast θ, but now m forecasts $f_1, f_2, \ldots, f_m$ are available. The m forecast errors $f_1 - \theta, \ldots, f_m - \theta$ are assumed to have a zero mean vector and covariance matrix $\mathbf{\Sigma}$. The combined forecast under this assumption is a weighted average

$$f_c(m) = \mathbf{u}'\mathbf{\Sigma}^{-1}\mathbf{f}/\mathbf{u}'\mathbf{\Sigma}^{-1}\mathbf{u} = \sum_{i=1}^{m} w_{im} f_i, \quad (4)$$

where $\mathbf{u} = (1, \ldots, 1)'$, $\mathbf{f} = (f_1, \ldots, f_m)'$, a prime denotes transposition, and w_{im} is the ith element of the vector $\mathbf{u}'\mathbf{\Sigma}^{-1}/\mathbf{u}'\mathbf{\Sigma}^{-1}\mathbf{u}$. The variance of the error $f_c(m) - \theta$ is $\sigma_c^2(m) = 1/\mathbf{u}'\mathbf{\Sigma}^{-1}\mathbf{u}$.

Instead of combining all m forecasts at once, we could combine them sequentially in $m - 1$ steps, adding one forecast at each step as in a stepwise regression with forward selection. This is more complicated than simply combining $f_1, \ldots, f_m$ in one fell swoop, but it means that at each step only two forecasts are being combined, the new forecast entering the combination and an aggregate of the forecasts already considered. Thus, we can utilize the graphs and the variability of the estimated weights from Section 1 to gain insight for the m-forecast case.

The sequential procedure consists of incorporating f_j after considering $f_1, \ldots, f_{j-1}$ for $j = 2, \ldots, m$. Adding another forecast f_j will change the weights given to $f_1, \ldots, f_{j-1}$. For $i = 1, \ldots, j - 1$, let $v_{ij} = w_{ij}/\sum_{i=1}^{j-1} w_{ij}$, and define

$$f_{1j}^{**} = \sum_{i=1}^{j-1} v_{ij} f_i \text{ and } f_{2j}^{**} = f_j.$$

Treating f_{1j}^{**} and f_{2j}^{**} as two forecasts and combining them via the procedures in Section 1 yields the same results as combining $f_1, \ldots, f_j$ using (4). Let σ_{ij}^{**} represent the standard deviation of $f_{ij}^{**} - \theta$ for $i = 1, 2$, ρ_j^{**} the correlation of $f_{1j}^{**} - \theta$ and $f_{2j}^{**} - \theta$, and ϕ_j^{**} the ratio $\sigma_{1j}^{**}/\sigma_{2j}^{**}$. The weights assigned to f_{1j}^{**} and f_{2j}^{**} when they are combined are $w_{1j}^{**} = 1 - w_{jj}$ and $w_{2j}^{**} = w_{jj}$. Now (1) applies to the sequential revision, with a second subscript added to the symbols to index the forecast being incorporated and a double asterisk to indicate the sequential combination process. Using this sequential procedure, we can see how adding forecasts to our combination influences the sensitivity of the combined forecast.

2.2. Exchangeable Forecasts

A special case occurs when we have equal standard deviations ($\sigma_i = \sigma$ for all i) and equal correlations ($\rho_{ij} = \rho$ for all $i \neq j$), implying that if these parameters are known, the forecasts receive equal weights. This case is of particular interest because we often view the information sources as exchangeable or nearly so. A decision maker may wish to consult experts or established models, and poor forecasters or models are presumably driven from the market. Furthermore, forecasters often have access to similar information and techniques.

Because the weights are always equal, $v_{ij} = w_{i,j-1} = 1/(j - 1)$. From Clemen and Winkler (1985),

$$\sigma_{1j}^{**} = \sigma\sqrt{[1 + (j - 2)\rho]/(j - 1)}.$$

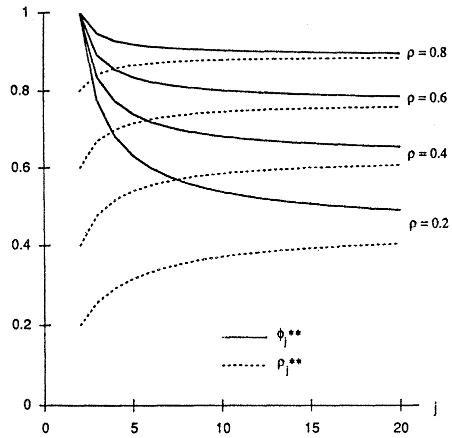

Figure 4. Movement of sequential pairs $(\phi_j^{**}, \rho_j^{**})$ as j increases for selected values of ρ.

Of course, $\sigma_{2j}^{**} = \sigma$, so that

$$\phi_{ij}^{**} = \sqrt{[1 + (j-2)\rho]/(j-1)}.$$

The covariance of $f_{1j}^{**} - \theta$ and $f_{2j}^{**} - \theta$ is $\sigma^2\rho$, and

$$\rho_j^{**} = \rho\sqrt{j-1}/\sqrt{1 + (j-2)\rho}.$$

What happens to ϕ_j^{**} and ρ_j^{**} as j increases? If $\rho > 0$, which appears to be the case in virtually every empirical study to date, then $d\rho_j^{**}/dj > 0$, with $\lim_{j\to\infty} \rho_j^{**} = \sqrt{\rho}$. Moreover, $d\phi_j^{**}/dj < 0$, with $\sigma_{12}^{**}/\sigma_{22}^{**} = 1$ and $\lim_{j\to\infty} \phi_j^{**} = \sqrt{\rho}$. Thus, as j increases, ρ_j^{**} and ϕ_j^{**} approach each other, making it more likely that sampling variability in the estimation of the model parameters will result in a negative weight.

The shifts as j increases in the sequential process are shown in Figure 4 for various values of ρ; note, in particular, the shifts for the higher values of ρ. For example, when $\rho = 0.8$, which is not at all extreme in view of many empirical studies, ϕ_j^{**} and ρ_j^{**} start relatively close and quickly move closer. In this case, we see that the sequence of $(\phi_j^{**}, \rho_j^{**})$ pairs approaches a point high on the line from (0, 0) to (1, 1) in Figure 2. This line separates the region with both weights positive from the region with the second weight negative, and points high on the line near (1, 1) are particularly sensitive in terms of both the weights and the error variance for the combined forecast (see Section 1).

As noted, the exchangeable case is a situation that may be reasonable in many realistic applications. Moreover, we might intuitively expect it to be relatively insensitive because the ideal weights are equal weights, which are far removed from scenarios with negative weights or extremely high weights. However, the behavior of ρ_j^{**} and ϕ_j^{**}, considered in light of the results from Section 1, indicates a surprising degree of sensitivity that can be a serious problem even in the exchangeable case.

2.3. Nonexchangeable Forecasts

With nonexchangeable forecasts, the decision maker is likely to consult those forecasters perceived as "better" first, suggesting that σ_{2j}^{**} will increase with j as σ_{1j}^{**} is decreasing with j. As a result, ϕ_j^{**} will decrease rapidly. At the same time, ρ_j^{**} might be expected to increase; as j gets larger, more information is contained in $f_1, \ldots, f_{j-1}$ and the degree of redundancy in f_j should increase. Also, a reasonable strategy would be to consult the "least redundant" forecasters early in the sequential process, contributing to the tendency of ρ_j^{**} to increase. Once again, then, we find ourselves moving toward the sensitive portion of Figure 2, and we might expect this movement to be more rapid in the nonexchangeable case than in the exchangeable case.

To illustrate the nonexchangeable case, in which general analytical results are difficult to obtain, we consider an example with four forecasters (or forecast methods). The variances of the forecast errors are $\sigma_1^2 = 1.00$, $\sigma_2^2 = 1.50$, $\sigma_3^2 = 1.65$, and $\sigma_4^2 = 1.90$; the correlations among forecast errors are $\rho_{12} = 0.50$, $\rho_{13} = \rho_{14} = \rho_{23} = 0.60$, $\rho_{24} = 0.70$, and $\rho_{34} = 0.95$. In this example, it is optimal (in the sense of minimizing σ_c^2 at each step) to bring the forecasts into the combination in reverse order of their variances, as in Table I.

Note that ϕ_j^{**} decreases as j increases, while ρ_j^{**} increases dramatically. By the time $j = 4$, ϕ_4^{**} is considerably smaller than ρ_4^{**}; as a result, w_{14}^{**} is greater than 1 and w_{24}^{**} is less than zero. This is in

Table I
Sequential Combination of Forecasts in Example

Step (j)	Forecast Added	ϕ_j^{**}	ρ_j^{**}	w_{1j}^{**}	w_{2j}^{**}	σ_c^2
1	f_1	—	—	—	—	1.000
2	f_2	0.816	0.500	0.696	0.304	0.882
3	f_3	0.731	0.681	0.932	0.068	0.878
4	f_4	0.723	0.916	1.708	−0.708	0.804

contrast to the exchangeable case, where ϕ_j^{**} approaches but is always greater than ρ_j^{**}. Referring once again to Figure 2, we see that the nonexchangeable case can quickly push us above and to the left of the line from (0, 0) to (1, 1), as it does in the example when $j = 4$. Furthermore, the point (0.723, 0.916) is where we would be if we knew all the parameters. We generally do not have this information, and when sampling fluctuations are considered, we could easily wind up with even more extreme weights.

3. DISCUSSION

The variability of weights has been discussed previously by ourselves and others, usually in terms of specific cases and speculation about the causes of poor performance of some procedures in empirical studies. The existence of a sensitivity problem is clearly understood. The contribution of this note is to add some precision to the discussion. The sampling distribution of the estimated weights and Figures 1–4 are helpful in understanding why and where the problem of the sensitivity of weights in combining forecasts arises. For example, Figure 2 indicates where in the parameter space sensitivity is likely to be a problem, and the standard deviation of the estimated weight graphed in Figure 3 shows how large the variability in weights can be. Furthermore, by looking at the combination of m forecasts in a sequential manner, we can use the graphs and sampling-distribution results to study sensitivity in the m-forecast case as well as in the two-forecast case.

The sensitivity shown here may be only the tip of the iceberg. Additional problems can arise if assumptions of the model or the data collection process are not met. For example, fat tails in the distributions of forecast errors might be especially likely to create difficulties. A more pervasive problem is nonstationarity. Even if the model is appropriate at any given time, parameters are likely to change over time; the underlying process could change, individual forecasting models or forecasters could adjust over time in response to previous errors, or changes in the degree of communication among forecasters and publishing of forecasts could modify the degree of dependence among forecasts. The results in this paper for examples with $n = 30$ are highly sensitive with high correlations, and nonstationarity could mean that even samples of this size with essentially the same parameters are unlikely to be available.

Combining forecasts has received considerable attention of late and appears to provide improved forecasts (Clemen 1989), Intuitively, combining forecasts should reduce the risk of an extremely bad forecast. If the combining technique is very sensitive and the weights fluctuate widely, this reduction in risk could be more than offset by an increase in risk associated with the possibility of extreme weights (particularly negative weights and weights greater than one). Our focus here is on a better understanding of the sensitivity problem. In particular, the development of new prescriptive methods to circumvent this problem is outside the scope of this note. Promising alternatives with reduced variability of weights include: using a simple average; basing weights on the estimated relative precision of the forecasts, thereby implicitly assuming zero correlations (Granger and Newbold 1977, Bunn 1985); using a Bayesian model to shrink the estimated weights toward a prior estimate, such as a simple average (Clemen and Winkler 1986); choosing which parameters to estimate when faced with correlated forecasts and limited data (Schmittlein, Kim and Morrison 1990); and using outperformance measures (Bunn 1975, 1977, Gupta and Wilton 1987).

ACKNOWLEDGMENT

This research was supported in part by the National Science Foundation under grants IST-8600788, SES-9022616, and SES-9022573.

REFERENCES

Bunn, D. W. 1975. A Bayesian Approach to the Linear Combination of Forecasts. *Opnl. Res. Quart.* **26,** 325–329.

Bunn, D. W. 1977. A Comparative Evaluation of the Outperformance and Minimum Variance Procedures for Linear Syntheses of Forecasts. *Opnl. Res. Quart.* **28,** 653–660.

Bunn, D. W. 1985. Statistical Efficiency in the Linear Combination of Forecasts. *Int. J. Forecasting* **1,** 151–163.

Clemen, R. T. 1989. Combining Forecasts: A Review and Annotated Bibliography. *Int. J. Forecasting* **5,** 559–583.

Clemen, R. T., and R. L. Winkler. 1985. Limits for the Precision and Value of Information From Dependent Sources. *Opns. Res.* **33,** 427–442.

Clemen, R. T., and R. L. Winkler. 1986. Combining Economic Forecasts. *J. Bus. Econ. Statist.* **4,** 39–46.

Granger, C. W. J., and P. Newbold. 1977. *Forecasting Economic Time Series.* Academic Press, London.

Gupta, S., and P. C. Wilton. 1987. Combination of Forecasts: An Extension. *Mgmt. Sci.* **33,** 356–372.

Kang, H. 1986. Unstable Weights in the Combination of Forecasts. *Mgmt. Sci.* **32,** 683–695.

KENDALL, M. G., AND A. STUART. 1963. *The Advanced Theory of Statistics*, Vol. 1, 2nd ed. Hafner, New York.

NEWBOLD, P., AND C. W. J. GRANGER. 1974. Experience With Forecasting Univariate Time Series and the Combination of Forecasts. *J. Roy. Statist. Soc.* A **137,** 131–149.

SCHMITTLEIN, D. C., J. KIM, AND D. G. MORRISON. 1990. Combining Forecasts: Operational Adjustments to Theoretically Optimal Rules. *Mgmt. Sci.* **36,** 1044–1056.

WINKLER, R. L. 1981. Combining Probability Distributions From Dependent Information Sources. *Mgmt. Sci.* **27,** 479–488.

THE EFFECT OF EXTERNALIZING SETUPS IN THE ECONOMIC LOT SCHEDULING PROBLEM

GUILLERMO GALLEGO and ILKYEONG MOON

Columbia University, New York, New York

(Received September 1989; revisions received March, December 1990; accepted March 1991)

This paper considers the Economic Lot Scheduling Problem with Reducible Setup Times (ELSP-RS), that is, determining a multiple product, single facility cyclic schedule to minimize holding and setup costs when setup times can be reduced, at the expense of setup costs, by externalizing setup operations. We develop an efficient algorithm that finds an optimal or near-optimal production schedule. Computational results indicate that dramatic savings are possible for highly utilized facilities.

It is generally assumed that setup times and setup costs are fixed. Here we will take a different view and attempt to explain how, in highly utilized facilities, reductions in average cost can be obtained by externalizing setup operations. Our discussion will be within the framework of the Economic Lot Scheduling Problem (ELSP), but it is applicable to capacitated production planning as well. In the ELSP, one facility is used to produce a variety of items. The items are produced in lots, and the facility is set up each time a lot for a different item is produced. The ELSP concerns the determination of the sequence and the sizes of the lots needed to minimize the long-run average setup and holding costs.

To present our ideas in a more precise way, we will separate the setup times into two parts: the *internal*—the part done while the facility is stopped; and the *external*—the part done while the facility is operating. As such, we will use setup time to mean internal setup time. Each setup is assigned a cost, hereby called the *setup cost*, which is the sum of all the out-of-pocket expenses associated with setup operations—both internal and external. The opportunity cost of the facility itself is, however, *not included* in the setup cost.

In their quest for just-in-time (JIT) manufacturing, many Japanese manufacturers have strived to reduce setup times in order to produce smaller lot sizes. They have often achieved setup time reduction through the externalization of some setup operations (see Shingo 1985). Reductions in setup times have an impact on the out-of-pocket setup costs, in that, in general, there is an *increase* in the out-of-pocket setup costs as the internal setup operations are externalized. This increase occurs because such operations are often more time consuming when done off-line—they require additional or better trained workers, or more careful coordination by management. However, traditional inventory theory clearly shows that larger setup costs result in larger economic lot sizes. The only possible justification for this paradoxical behavior where setup costs are increased in order to reduce lot sizes appears to lie in other benefits derived from JIT, for example, faster customer response time, better quality control, increased production flexibility, and smaller space requirements.

We assert, however, that this need not be the case when there is a production capacity constraint linking the items. This production capacity constraint states that the proportion of time available for setups and

Subject classifications: Production/scheduling: economic lot sizing problem. Production/scheduling, approximations/heuristic: setup time reduction. *Area of review:* MANUFACTURING, PRODUCTION AND SCHEDULING.