



Limits for the Precision and Value of Information from Dependent Sources

Robert T. Clemen, Robert L. Winkler

Operations Research, Volume 33, Issue 2 (Mar. - Apr., 1985), 427-442.

Stable URL:

<http://links.jstor.org/sici?sici=0030-364X%28198503%2F198504%2933%3C427%3ALFTPAV%3E2.0.CO%3B>

Your use of the JSTOR archive indicates your acceptance of JSTOR's Terms and Conditions of Use, available at <http://www.jstor.org/about/terms.html>. JSTOR's Terms and Conditions of Use provides, in part, that unless you have obtained prior permission, you may not download an entire issue of a journal or multiple copies of articles, and you may use content in the JSTOR archive only for your personal, non-commercial use.

Each copy of any part of a JSTOR transmission must contain the same copyright notice that appears on the screen or printed page of such transmission.

Operations Research is published by INFORMS. Please contact the publisher for further permissions regarding the use of this work. Publisher contact information may be obtained at <http://www.jstor.org/journals/informs.html>.

Operations Research
©1985 INFORMS

JSTOR and the JSTOR logo are trademarks of JSTOR, and are Registered in the U.S. Patent and Trademark Office. For more information on JSTOR contact jstor-info@umich.edu.

©2002 JSTOR

Limits for the Precision and Value of Information from Dependent Sources

ROBERT T. CLEMEN

University of Oregon, Eugene, Oregon

ROBERT L. WINKLER

Duke University, Durham, North Carolina

(Received June 1983; accepted May 1984)

In many inferential and decision-making situations, information is obtained from a number of information sources, and the separate pieces of information are often not independent. This paper investigates the impact of dependence on the precision and value of information. The results indicate that positive dependence among information sources can have a serious detrimental effect on the precision and value of the information. Differences in precision between the dependent and independent cases can be remarkably large. With dependence, the incremental value of information can decrease very rapidly, and the limiting value of information as more sources are considered can be considerably less than the expected value of perfect information. The results of this paper have implications for the acquisition and use of information in decision-making problems.

IMPORTANT real-world situations invariably involve uncertainties about future events or variables. Inferences and decisions must be made in the face of such uncertainties, and information that helps to reduce the uncertainty can play a crucial role in the inferential and decision-making processes. After obtaining additional information, a decision maker might be able to make more precise estimates of random variables and might be able to reduce the risk associated with potential actions that are being considered for implementation.

In many informal and formal inferential and decision-making situations, information is obtained from a number of information sources. Examples are subjective judgments from various experts and forecasts from different forecasting methods (e.g., exponential smoothing, Box-Jenkins) or models (e.g., econometric models). The notion that more information leads to better inferences and decisions motivates the consideration of multiple information sources.

Often, we would expect information from different sources to be somewhat redundant. For example, the different sources might utilize

Subject classification: 95 value of information, 793 revision with dependent information.

some common data, share common assumptions, or have access to some of each other's opinions. We can represent such redundancy stochastically in terms of positive dependence among the information sources.

Intuitively, it seems that positive dependence among information sources should reduce the overall information content and the gain in precision afforded by the information. This potential reduction in precision could decrease the expected value of the information. In this paper, we demonstrate that changes in precision and value as a result of positive dependence do in fact generally move in the anticipated direction and that the magnitude of such changes can be substantial. These results have important normative implications for the acquisition and use of information in inferential and decision-making problems.

The purpose of this paper, then, is to study the impact of dependence on the precision and value of information coming from a number of sources. Relatively few efforts have dealt with the acquisition and use of information from dependent sources. Here, we study the impact of dependence among information sources in the context of a model based on Winkler [1981]. For some other work involving dependent information sources, see Geisser [1965], Hogarth [1977], Morris [1977], Bunn [1978], and French [1980].

The paper is organized as follows. First, we present the model, which is based on information sources that provide estimates with a joint distribution of estimation errors. The case in which the joint distribution is multivariate normal receives special attention because it is tractable and should often provide a reasonable approximation in practice.

Next, in Section 2, we study the inferential aspects of the problem. We develop the notion that a set of dependent information sources produces information that, in the aggregate, is equivalent to that produced by a smaller number of independent sources. Bounds and limiting results are obtained for the precision of information from dependent sources, and the upper bounds tend to be remarkably low.

In Section 3 we turn to decision making and study the expected value of information from dependent sources. Some general results indicate that information from a set of dependent sources can be expected to be less valuable than information from independent sources with the same individual precisions. In the context of a special class of decision-making problems, we investigate the magnitude of the reduction in value of information associated with dependence and relate this reduction to the number of sources and the amount of prior information available to the decision maker. The limiting value of information as more sources are considered is generally considerably less than the expected value of perfect information.

Section 4 contains a numerical example. Section 5 discusses situations in which some assumptions used in Sections 2 and 3 are violated and

dependence may actually be helpful. A summary and a brief discussion of some implications of the results concludes the paper in Section 6.

1. A MODEL WITH DEPENDENT INFORMATION SOURCES

Let the variable of interest, which could be either a future observation or a parameter of a statistical model, be denoted by θ . An individual, called the decision maker, is interested in θ because θ may affect the ultimate consequence of a decision-making problem. The decision maker has a prior density $h_0(\theta)$ for θ and can obtain additional information about θ . In particular, the decision maker has access to k information sources, each of which provides a point estimate of θ . For the purpose of this paper, the sources of information will all be called experts, although in practice some sources may be data sets or empirical or theoretical models.

If consulted by the decision maker, expert i ($i = 1, \dots, k$) will provide a point estimate x_i of θ . The error of estimation for expert i is

$$u_i = x_i - \theta,$$

and dependence among the experts will be represented by dependence among their estimation errors. If all k experts are consulted, the likelihood function can be taken as the joint density $f(\mathbf{u} | \theta)$, where $\mathbf{u} = (u_1, \dots, u_k)^t$ and t denotes transposition. The posterior density of θ is then

$$h(\theta | \mathbf{x}) \propto h_0(\theta) f(\mathbf{u} | \theta), \quad (1)$$

where $\mathbf{x} = (x_1, \dots, x_k)^t$. Moreover, if some subset of the k experts is consulted, the revision is of the same form as (1), with \mathbf{x} and \mathbf{u} replaced by the appropriate subsets. The density for any subset of \mathbf{u} can be found from $f(\mathbf{u} | \theta)$, so that the specification of this overall joint likelihood function is sufficient to allow consideration of all possible combinations of information sources.

This model is basically the model developed in Winkler. We assume here that the experts provide point estimates, not entire distributions for θ . However, an alternative way of interpreting the model is to assume that entire distributions are assessed and that the marginal distributions $f(u_i | \theta)$ from $f(\mathbf{u} | \theta)$ for $i = 1, \dots, k$ are based on the experts' assessed distributions (possibly calibrated by the decision maker) for θ . For more details, see Winkler.

The "normal theory of errors" suggests that a normal model for $f(\mathbf{u} | \theta)$ may be reasonable in many instances, and such a model is much more tractable than other multivariate models. Hence, the main results in this paper will be generated under the assumption that $f(\mathbf{u} | \theta)$ is normal with a zero mean vector and a positive definite covariance matrix Σ . The zero

mean vector implies that the experts are calibrated or unbiased; we can treat violations by calibrating the experts (i.e., by adjusting their estimates to provide expected estimation errors of zero). However, such calibration, which might be based on empirical evidence regarding past errors of estimation or on subjective judgments (see Morris), may be complicated by potential dependence among experts in terms of degree of noncalibration. Methods for estimating the covariance matrix Σ from empirical evidence are discussed in Bates and Granger [1969], Newbold and Granger [1974], and Winkler and Makridakis [1983], and the subjective assessment of correlation coefficients is studied by Gokhale and Press [1982].

Suppose that the decision maker's prior distribution for θ is normal with mean m_0 and variance σ_0^2 . We assume that the "prior estimation error" $m_0 - \theta$ is uncorrelated with any of the experts' errors. (A nonzero correlation between $m_0 - \theta$ and any u_i could be handled by treating the problem as one with a diffuse prior distribution and $k + 1$ experts, the decision maker being one of the experts.)

If the decision maker consults k experts, the posterior distribution of θ is a normal distribution with mean m_* and variance σ_*^2 , where

$$m_* = (\sigma_0^{-2}m_0 + \mathbf{e}^t \Sigma^{-1} \mathbf{x}) \sigma_*^2, \quad (2)$$

$$\sigma_*^2 = (\sigma_0^{-2} + \mathbf{e}^t \Sigma^{-1} \mathbf{e})^{-1}, \quad (3)$$

and $\mathbf{e}^t = (1, \dots, 1)$. The posterior mean m_* is a weighted average of the prior mean and the experts' estimates, with the weights depending on the covariance matrix. The posterior variance, and hence the gain in precision $\sigma_*^{-2} - \sigma_0^{-2}$ as a result of consulting the k experts, also depend on the covariance matrix. In the next section, we investigate the impact of dependence among the experts on potential gains in precision.

2. BOUNDS ON THE PRECISION OF INFORMATION FROM DEPENDENT SOURCES

When information sources are dependent, we can expect that the resulting information from these sources will be somewhat redundant. As a result, the decision maker may find that the reduction in variance or gain in precision afforded by consulting a set of dependent experts will be less than could be obtained if the experts were independent. To study this phenomenon, we first introduce the notion of an equivalent number of independent experts and then develop some limiting results and bounds for this measure.

Equivalent Number of Independent Experts

Suppose that the decision maker had the option of consulting n independent experts, each with error variance σ^2 . We can express the

decision maker's prior variance σ_0^2 in the form σ^2/n_0 , where n_0 can be thought of as an equivalent sample size for the purpose of assessing or interpreting a prior distribution (Raiffa and Schlaifer [1961]). Then the posterior variance following the consultation of the n independent experts would be

$$\sigma_*^2 = \sigma^2/(n_0 + n) = (\sigma_0^{-2} + n\sigma^{-2})^{-1}. \tag{4}$$

In the model in Section 1, the experts are not, in general, independent. With a multinormal distribution of estimation errors having covariance matrix Σ , the posterior variance is given by (3). The value of n that gives the same posterior variance with n independent experts having error variance σ^2 as with k experts having covariance matrix Σ will be called the equivalent number of independent experts with error variance σ^2 and will be denoted by $n(\sigma^2, \Sigma)$.

Comparing (3) and (4), we see that the two posterior variances are equal if $n\sigma^{-2} = \mathbf{e}^t \Sigma^{-1} \mathbf{e}$. Thus, the equivalent number of independent experts is

$$n(\sigma^2, \Sigma) = \sigma^2 \mathbf{e}^t \Sigma^{-1} \mathbf{e}. \tag{5}$$

The role of the measure $n(\sigma^2, \Sigma)$ is to help us interpret the contribution of the information from the k dependent experts. In using this measure, we will try to choose σ^2 so that the difference between $n(\sigma^2, \Sigma)$ and k will reflect primarily the dependence among experts.

Common Correlation, Common Variance

A special case that is tractable and provides some "feel" for the impact of dependence is a situation where Σ is an intraclass correlation matrix. In such a matrix, all variances are equal and all correlations are equal. This situation would arise, for example, when the experts are judged to be perfectly exchangeable.

Let the common variance and correlation be denoted by σ^2 and ρ , respectively, with $\rho > 0$. (We will discuss the case with $\rho < 0$ in Section 5.) From Press [1972], we can express Σ^{-1} in a relatively simple form, and

$$\mathbf{e}^t \Sigma^{-1} \mathbf{e} = k\sigma^{-2}[1 + (k - 1)\rho]^{-1}. \tag{6}$$

Applying (5) yields

$$n(\sigma^2, \Sigma) = k[1 + (k - 1)\rho]^{-1}. \tag{7}$$

Here, the n independent and k dependent experts all have the same error variance σ^2 . When $k > 1$, $n(\sigma^2, \Sigma) < k$; positive dependence reduces the information content of a set of experts' estimates. Moreover, the stronger the dependence, the greater the reduction, since $dn(\sigma^2, \Sigma)/d\rho < 0$.

The equivalent number of independent experts is a concave function of k with

$$\lim_{k \rightarrow \infty} n(\sigma^2, \Sigma) = \rho^{-1}.$$

Thus, there is an upper limit on the number of equivalent independent experts and on the precision of the information that can be attained by consulting exchangeable, but dependent experts. The limit is quite low: if $\rho = 0.80$, a value consistent with some empirical results we have obtained for football predictions (Winkler) and economic forecasts, then $n(\sigma^2, \Sigma) = 1.25$. Consequently, after the first expert, who is worth one complete independent expert, all other experts combined are worth only one-quarter of an independent expert. Even if $\rho = 0.25$, which is not very high, the best the decision maker could do would be to attain an equivalent of four independent experts. Figure 1 shows the behavior of $n(\sigma^2, \Sigma)$ as k varies for different values of ρ . After the first expert, the equivalent value in independent experts of each additional dependent expert is positive, but diminishes rapidly. This marginal equivalent value for the k th expert is

$$\begin{aligned} \Delta n(k) &= n(\sigma^2, \Sigma | k) - n(\sigma^2, \Sigma | k - 1) \\ &= (1 - \rho)[1 + (k - 1)\rho]^{-1}[1 + (k - 2)\rho]^{-1}, \end{aligned}$$

which is a decreasing convex function of k .

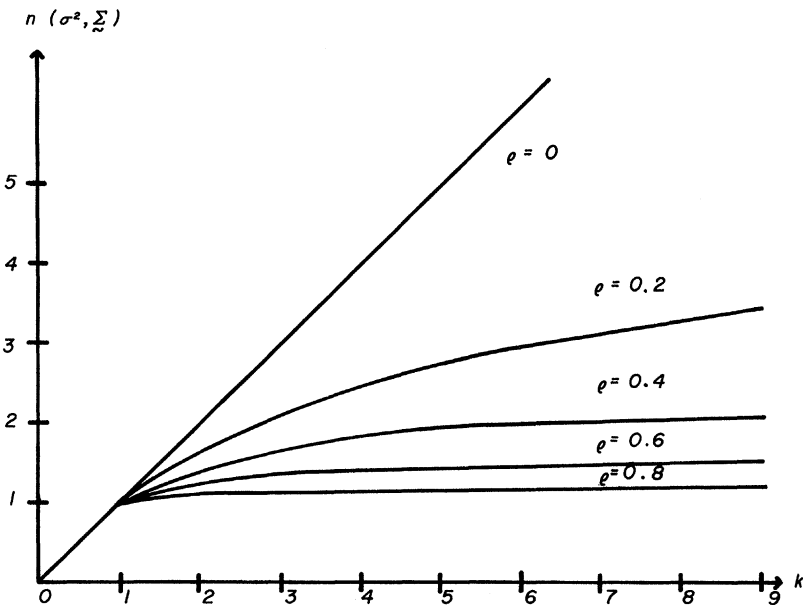


Figure 1. The equivalent number of independent experts as a function of k , the actual number of experts, for selected values of ρ .

Common Correlation, Different Variances

Here, we retain the assumption of a common correlation $\rho > 0$, but allow the error variances to differ. Applying the Binomial Inverse Theorem (Press), we find that

$$\mathbf{e}^t \boldsymbol{\Sigma}^{-1} \mathbf{e} = \frac{[1 + (k - 1)\rho] \sum_{j=1}^k \sigma_j^{-2} - \rho(\sum_{j=1}^k \sigma_j^{-1})^2}{(1 - \rho)[1 + (k - 1)\rho]} \tag{8}$$

To guarantee that all estimates will be given positive weights in using (2) to compute the posterior mean (Winkler), we assume that

$$\rho < [1 - k + \sigma_i \sum_{j=1}^k \sigma_j^{-1}]^{-1} \tag{9}$$

for $i = 1, \dots, k$. (We will briefly discuss the case of negative weights in Section 5.)

All other things being equal, a decision maker should prefer experts with smaller error variances, which is the case here, since, for $i = 1, \dots, k$,

$$\frac{d\mathbf{e}^t \boldsymbol{\Sigma}^{-1} \mathbf{e}}{d\sigma_i^{-1}} = \frac{2\{[1 + (k - 1)\rho]\sigma_i^{-1} - \rho \sum_{j=1}^k \sigma_j^{-1}\}}{(1 - \rho)[1 + (k - 1)\rho]} > 0 \tag{10}$$

as long as (9) is satisfied. From (10) and (3), a smaller σ_i implies a larger $\mathbf{e}^t \boldsymbol{\Sigma}^{-1} \mathbf{e}$, which in turn implies a smaller posterior variance σ_*^2 . As a result, from (5), if $k > 1$ and $\sigma_i \neq \sigma_j$ for some i and j ,

$$n(\sigma^2, \boldsymbol{\Sigma}_M) < n(\sigma^2, \boldsymbol{\Sigma}) < n(\sigma^2, \boldsymbol{\Sigma}_m), \tag{11}$$

where $\boldsymbol{\Sigma}_M$ and $\boldsymbol{\Sigma}_m$ have intraclass correlation structure with correlation ρ and respective variances $\sigma_M^2 = \max_j \{\sigma_j^2\}$ and $\sigma_m^2 = \min_j \{\sigma_j^2\}$.

If only one expert is consulted, the decision maker should use the expert with the smallest variance, σ_m^2 . Thus, we will use $\sigma^2 = \sigma_m^2$ as our “process variance” for comparisons of equivalent sample sizes. From (5), (6), and (11) with $\sigma^2 = \sigma_m^2$,

$$(\sigma_m^2 / \sigma_M^2) k [1 + (k - 1)\rho]^{-1} < n(\sigma_m^2, \boldsymbol{\Sigma}) < k [1 + (k - 1)\rho]^{-1}. \tag{12}$$

To obtain an even tighter lower bound for a fixed k , we consider $\sigma_p^2 = (\sum_{j=1}^k \sigma_j^{-2} / k)^{-1}$, the reciprocal of the average precision, where expert j 's precision is σ_j^{-2} . Letting $\boldsymbol{\Sigma}_p$ have intraclass correlation structure, with correlation ρ and variance σ_p^2 , we have, from (8) and (6),

$$\begin{aligned} \mathbf{e}^t \boldsymbol{\Sigma}^{-1} \mathbf{e} - \mathbf{e}^t \boldsymbol{\Sigma}_p^{-1} \mathbf{e} &= \frac{[1 + (k - 1)\rho] \sum_{j=1}^k \sigma_j^{-2} - \rho(\sum_{j=1}^k \sigma_j^{-1})^2}{(1 - \rho)[1 + (k - 1)\rho]} - \frac{\sum_{j=1}^k \sigma_j^{-2}}{1 + (k - 1)\rho} \\ &= \rho [k \sum_{j=1}^k \sigma_j^{-2} - (\sum_{j=1}^k \sigma_j^{-1})^2] / (1 - \rho)[1 + (k - 1)\rho] > 0. \end{aligned}$$

Therefore,

$$(\sigma_m^2/\sigma_p^2)k[1 + (k - 1)\rho]^{-1} < n(\sigma_m^2, \mathbf{\Sigma}) < k[1 + (k - 1)\rho]^{-1}.$$

With a starting point of equal variances σ_m^2 , the effect of increasing one or more variances is to deflate the equivalent sample size by a factor less extreme than the ratio σ_m^2/σ_p^2 .

General Covariance Matrix

Now we are prepared to consider the general case in which not only the error variances σ_i^2 , but also the correlations ρ_{ij} between errors, may differ. To assure positive weights in using (2) to compute the posterior mean, we assume that

$$\alpha_i = (\mathbf{e}^t \mathbf{\Sigma}^{-1})_i > 0 \quad \text{for } i = 1, \dots, k. \tag{13}$$

The impact on $\mathbf{e}^t \mathbf{\Sigma}^{-1} \mathbf{e}$ of changing one covariance σ_{ij} by changing the correlation coefficient ρ_{ij} while holding everything else constant is

$$d\mathbf{e}^t \mathbf{\Sigma}^{-1} \mathbf{e} / d\rho_{ij} = (d\mathbf{e}^t \mathbf{\Sigma}^{-1} \mathbf{e} / d\sigma_{ij})(d\sigma_{ij} / d\rho_{ij}) = \sigma_i \sigma_j (d\mathbf{e}^t \mathbf{\Sigma}^{-1} \mathbf{e} / d\sigma_{ij}). \tag{14}$$

But from Graybill [1969], if $\Delta_{ij} = (\delta_{rs})$ with $r, s = 1, \dots, k$ and

$$\delta_{rs} = \begin{cases} 1 & \text{if } r = i \text{ and } s = j \text{ or } r = j \text{ and } s = i, \\ 0 & \text{otherwise,} \end{cases}$$

then

$$d\mathbf{e}^t \mathbf{\Sigma}^{-1} \mathbf{e} / d\sigma_{ij} = -\mathbf{e}^t \mathbf{\Sigma}^{-1} \Delta_{ij} \mathbf{\Sigma}^{-1} \mathbf{e} = -2\alpha_i \alpha_j. \tag{15}$$

Therefore, combining (14) and (15) and using (13),

$$d\mathbf{e}^t \mathbf{\Sigma}^{-1} \mathbf{e} / d\rho_{ij} = -2\alpha_i \alpha_j \sigma_i \sigma_j < 0. \tag{16}$$

Condition (16) implies that, all other things being equal, a decision maker should prefer experts with smaller correlations between errors. Suppose that $\mathbf{\Sigma}_R(\mathbf{\Sigma}_r, \mathbf{\Sigma}_0)$ has common correlation $\rho_R = \max_{i,j} \{\rho_{ij}\}$ ($\rho_r = \min_{i,j} \{\rho_{ij}\}$, $\rho_0 = 0$) and variances identical to those in $\mathbf{\Sigma}$. Then, if $\rho_r > 0$ and $\rho_{ij} \neq \rho_{pq}$ for some i, j, p , and q ,

$$n(\sigma^2, \mathbf{\Sigma}_R) < n(\sigma^2, \mathbf{\Sigma}) < n(\sigma^2, \mathbf{\Sigma}_r) < n(\sigma^2, \mathbf{\Sigma}_0), \tag{17}$$

with the lower bound holding if ρ_R satisfies (9). Thus, a set of k dependent information sources with covariance matrix $\mathbf{\Sigma}$ will be less informative than k independent information sources having the same variances as in $\mathbf{\Sigma}$.

Using (11) to take (17) one step further gives us

$$n(\sigma^2, \mathbf{\Sigma}_{MR}) < n(\sigma^2, \mathbf{\Sigma}_R) < n(\sigma^2, \mathbf{\Sigma}) < n(\sigma^2, \mathbf{\Sigma}_r) < n(\sigma^2, \mathbf{\Sigma}_{mr}), \tag{18}$$

where $\mathbf{\Sigma}_{MR}$ and $\mathbf{\Sigma}_{mr}$ have an intraclass correlation structure, with respec-

tive correlations ρ_R and ρ_r and variances σ_M^2 and σ_m^2 . With σ_m^2 as our "process variance" for comparing equivalent sample sizes, as in (12), we have

$$(\sigma_m^2/\sigma_M^2)k[1 + (k - 1)\rho_R]^{-1} < n(\sigma_m^2, \Sigma) < k[1 + (k - 1)\rho_r]^{-1}. \quad (19)$$

As a corollary of (19), the equivalent number of independent experts is less than the actual number of dependent experts:

$$n(\sigma_m^2, \Sigma) < k.$$

After the first expert, dependence and possibly higher individual error variances reduce the precision as compared with the same number of independent experts exchangeable with the first expert.

3. THE VALUE OF DEPENDENT INFORMATION SOURCES IN DECISION MAKING

The results in Section 2 involve the inferential impact of dependent information sources. As noted in Section 1, however, the decision maker is interested in θ because θ may affect the ultimate consequence of a decision-making problem. Thus, the measure of primary concern to the decision maker in contemplating the purchase of information is the expected value of the information in terms of the decision-making problem. We first present some general comments concerning the value of information, and then investigate the implications of dependence in the context of a special type of decision-making problem.

Expected Value of Information

We define the expected value of information as the maximum amount that the decision maker could pay for the information without being worse off than he was without the information (e.g., LaValle [1968]). Suppose that the decision maker must choose an action $a \in A$ and that $c(a, \theta)$ represents the consequence received if a is chosen and θ occurs. The consequence $c(a, \theta)$ could be multidimensional, but is exclusive of any costs of information. Let $U[c(a, \theta) v]$ represent the decision maker's utility if $c(a, \theta)$ is received and information costs of v are incurred. Then, the expected value of obtaining information y , as calculated before y is actually obtained, is the value of v that solves

$$E_y \{ \max_{a \in A} E_{\theta|y} U[c(a, \theta) v] \} = \max_{a \in A} E_{\theta} U[c(a, \theta), 0].$$

Let $v(y)$ represent the expected value of information y . As long as the decision-making problem has no game-theoretic aspects (e.g., θ does not represent the action of an opponent who may change θ upon learning that the decision maker has purchased information), $v(y) \geq 0$. On an incremental basis, using Blackwell's [1953] notion of sufficiency,

$(x_1, \dots, x_k, x_{k+1})$ is sufficient for (x_1, \dots, x_k) for $k = 1, 2, \dots$. Thus,

$$v(x_1, \dots, x_k, x_{k+1}) \geq v(x_1, \dots, x_k).$$

This condition says that consulting an extra expert can never be detrimental in a value-of-information sense, regardless of whether the experts are dependent or independent. Of course, when the costs of consulting the experts are considered, consulting k experts might turn out to be preferable to consulting $k + 1$ experts.

With the normal model studied in Sections 1 and 2, consulting k experts with covariance matrix Σ yields the same posterior variance and the same prior distribution of m_* as consulting $n(\sigma^2, \Sigma)$ independent experts with error variance σ^2 . Thus, if $v(k, \Sigma)$ and $v_I[n(\sigma^2, \Sigma), \sigma^2]$, respectively, denote the expected values of the information from these two sets of information, then

$$v(k, \Sigma) = v_I[n(\sigma^2, \Sigma), \sigma^2]. \quad (20)$$

But, since extra information can never be detrimental, $v_I(n, \sigma^2)$ must be a nondecreasing function of n . We can therefore use (20) to express the ordering in (18) concerning precision as an ordering in terms of value of information:

$$v(k, \Sigma_{MR}) \leq v(k, \Sigma_R) \leq v(k, \Sigma) \leq v(k, \Sigma_r) \leq v(k, \Sigma_{mr}).$$

With (13) satisfied and all other things being equal, reducing an error variance or a correlation leads to the same or an improved value of information. Also, from (17) and (20),

$$v(k, \Sigma) \leq v(k, \Sigma_0),$$

which means that a set of independent experts with error variances $\sigma_1^2, \dots, \sigma_k^2$ is at least as valuable as a dependent set with the same variances and positive correlations.

Value of Information for the Newsboy Problem

It is difficult to make general statements about the value of information that are valid for any decision maker and decision-making problem (e.g., see Hilton [1981]). To get some idea of possible magnitudes of the impact of dependence on the value of information, we analyze a special, common type of decision-making problem, an infinite-action problem with linear loss. This situation is sometimes called the newsboy problem because it can be viewed as a choice by a newsboy of how many papers to order to minimize expected loss, given a probability distribution for demand (θ) and a linear loss function. It can also be viewed as a point estimation problem, with the action a representing an estimate of θ . The optimal

strategy for the newsboy is to set $a = F_{\theta}^{-1}(\kappa)$, where F_{θ} represents the decision maker's cumulative distribution function for θ , $\kappa = \kappa_u/(\kappa_u + \kappa_o)$, and κ_u and κ_o are the per unit costs of underestimation and overestimation (Raiffa and Schlaifer).

For the normal model, the optimal strategy under the prior distribution is $a = m_0 + z_{\kappa} \sigma_0$, where $z_{\kappa} = \Phi^{-1}(\kappa)$ and Φ represents the standard normal cumulative function. The corresponding expected loss is $(\kappa_u + \kappa_o) \sigma_0 \phi(\kappa)$, where ϕ denotes the standard normal density function. After information is obtained from k experts and the decision maker's distribution for θ is revised as in Section 2, the optimal strategy will be $a = m_* + z_{\kappa} \sigma_*$ with expected loss $(\kappa_u + \kappa_o) \sigma_* \phi(\kappa)$. The expected value of the information is

$$v(k, \Sigma) = (\sigma_0 - \sigma_*) (\kappa_u + \kappa_o) \phi(\kappa).$$

Suppose that Σ has intraclass correlation structure with common correlation $\rho > 0$ and variance σ^2 , and define $n_0 = \sigma^2/\sigma_0^2$ as the equivalent number of experts represented by the decision maker's prior distribution. Then

$$v(k, \Sigma) = \sigma(\kappa_u + \kappa_o) \phi(\kappa) (n_0^{-1/2} - \{n_0 + k[1 + (k - 1)\rho]\}^{-1/2}),$$

which is a decreasing function of ρ and n_0 and a concave increasing function of k . If σ_0^2 is fixed so that n_0 is a function of σ^2 , then $v(k, \Sigma)$ is a decreasing function of σ . The value of information increases as more experts are consulted, as the dependence between experts is reduced, as the error variance for individual experts decreases, and as the amount of prior information decreases.

To illustrate how $v(k, \Sigma)$ is affected by k , n_0 and ρ , we graph $v(k, \Sigma)/\sigma(\kappa_u + \kappa_o) \phi(\kappa)$ in Figure 2 as a function of k for selected values of n_0 and ρ . The proportional decrease in $v(k, \Sigma)$ as ρ increases for a fixed k and n_0 is greater as n_0 and k increase. The limiting value of $v(k, \Sigma)$ as $k \rightarrow \infty$ is also affected by n_0 and ρ :

$$\lim_{k \rightarrow \infty} v(k, \Sigma) = \sigma(\kappa_u + \kappa_o) \phi(\kappa) [n_0^{-1/2} - (n_0 + \rho^{-1})^{-1/2}].$$

In the limit, $v(k, \Sigma)$ is less than the value of perfect information about θ ,

$$v_I(\infty, \sigma^2) = \sigma(\kappa_u + \kappa_o) \phi(\kappa) n_0^{-1/2},$$

by a factor of $1 - [n_0/(n_0 + \rho^{-1})]^{1/2}$. This factor is decreasing in ρ and n_0 and is strictly less than one as long as $\rho > 0$. For example, if $n_0 = 0.5$ (relatively limited prior information) and $\rho = 0.6$ (moderate correlation), the value of an infinite number of dependent experts is $0.52v_I(\infty, \sigma^2)$. Positive dependence prevents $v(k, \Sigma)$ from approaching the value of perfect information no matter how large k becomes.

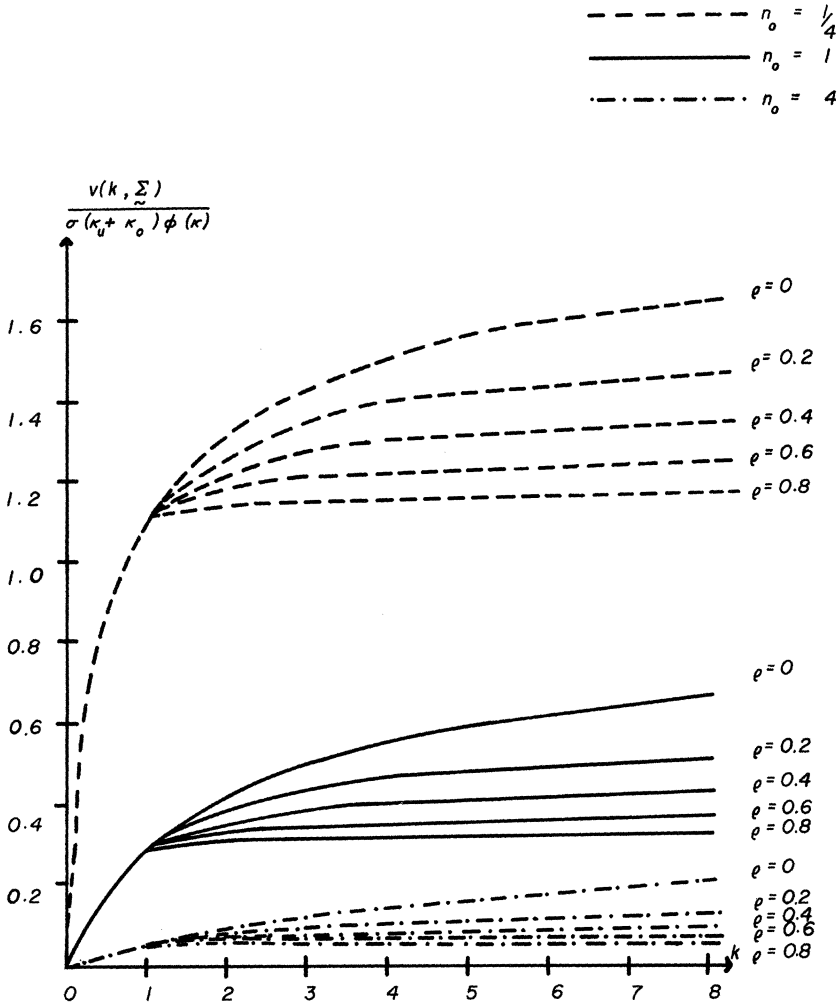


Figure 2. The value of information, divided by $\sigma(\kappa_u + \kappa_o)\phi(\kappa)$, as a function of k for selected values of n_o and ρ .

4. AN EXAMPLE

Suppose that a decision maker with a newsboy problem has very little prior information about θ and represents this state of information using a normal prior distribution with $\sigma_0^2 = 100$. Three experts are available, and the joint distribution of their estimation errors is normal with zero mean vector, $\sigma_1^2 = \sigma_2^2 = 10$, $\sigma_3^2 = 12$, $\rho_{12} = 0.6$, and $\rho_{13} = \rho_{23} = 0.37$.

For each possible combination of experts that can be consulted, Table I gives the equivalent number of independent experts (with $\sigma^2 = 10$), the posterior variance, and $v(k, \Sigma)/(\kappa_u + \kappa_o)\phi(\kappa)$. Expert 3, with a higher

variance than the other two experts, is equivalent only to 0.83 independent experts. However, if two experts are to be consulted, the combination of 3 with either 1 or 2 is worth 1.35 independent experts, as compared with only 1.25 independent experts for 1 and 2. A higher $n(\sigma^2, \Sigma)$, of course, translates into a lower posterior variance and an increased value of information, as illustrated in Table I. In this example, the higher variance of Expert 3 is offset by lower correlation ($\rho_{13} = \rho_{23} < \rho_{12}$).

Alternatively, suppose that Expert 1 was the decision maker. Then, if either Expert 2 or Expert 3 could be consulted, 3 would be preferred because of the variance/correlation tradeoff. However, since the experts are now correlated with the decision maker and the decision maker's own variance is 10 instead of 100, the value of information is much lower than the values given in Table I. Now $v(k, \Sigma)/(\kappa_u + \kappa_o)\phi(\kappa)$ is only 0.33 for Expert 2, 0.44 for Expert 3, and 0.57 for the combination of Experts 2 and 3. In contrast, if all correlations were zero, then Expert 2 would be preferred to Expert 3 and $v(k, \Sigma)/(\kappa_u + \kappa_o)\phi(\kappa)$ would be 1.27 for the combination of Experts 2 and 3.

5. CAN DEPENDENCE EVER BE HELPFUL?

We have shown that dependence can significantly reduce the precision and value of information from a number of sources. In deriving these results, we have placed certain restrictions on the nature of the dependence, allowing only positive correlations in the common correlation case and placing upper bounds on correlations to avoid the use of negative weights in the computation of the posterior mean. This section briefly discusses some implications of relaxing these constraints.

First, we consider the possibility of negative correlations. In the case of a common correlation and a common variance, suppose that $-k^{-1} \leq \rho < 0$ (the lower bound is necessary to make Σ positive definite). From (7), then,

$$k < n(\sigma^2, \Sigma) \leq k^2$$

TABLE I
PRECISION AND VALUE OF INFORMATION FROM AVAILABLE COMBINATIONS OF EXPERTS

Experts Consulted	$n(10, \Sigma)$	σ_*^2	$v(k, \Sigma)/(\kappa_u + \kappa_o)\phi(\kappa)$
1	1.00	9.09	6.98
2	1.00	9.09	6.98
3	0.83	10.71	6.73
1, 2	1.25	7.41	7.28
1, 3	1.35	6.89	7.38
2, 3	1.35	6.89	7.38
1, 2, 3	1.50	6.28	7.49

when $k > 1$. Since results from this case (with $\rho > 0$) are used to generate bounds on $n(\sigma^2, \Sigma)$ for more general cases in Section 2, such bounds might not be valid when negative correlations are allowed. Thus, negative dependence can lead to increases in the precision (and hence in the value) of information.

Next, we consider situations with highly positive correlations and some differences among error variances. In such situations, (13) (or (9), a special case of (13)) could be violated, resulting in negative weights for one or more experts when (2) is used to combine the experts' estimates. Therefore, the posterior mean may actually shift away from the estimates of some experts with high error variances. The implication for the results in Sections 2 and 3 is that various inequalities may be reversed and the equivalent number of independent experts may be less than, equal to, or greater than k .

Thus, it is possible for dependence among experts to be helpful in that it could increase the precision and value of information when correlations are negative or highly positive. However, although negative correlations can be helpful, in practice correlations among experts are generally positive. Highly positive correlations are not uncommon in empirical estimates of Σ (e.g., Figlewski [1983], Figlewski and Urich [1983]), but a violation of (13) in an estimate of Σ indicates an ill-conditioned matrix that suggests a multicollinearity problem. As a result, estimates of weights and equivalent sample sizes can be highly unstable. Therefore, although extremely high correlations can, in theory, lead to increased precision when compared with the independent case, estimation problems make it difficult, if not impossible, to realize this increased precision in practice.

6. SUMMARY AND DISCUSSION

The decision to obtain information from several sources is usually motivated by a desire to reduce uncertainty and thus to increase precision. In decision-making problems under uncertainty, the desire to increase precision is related to the expectation that greater precision will generally lead to improved, more informed decisions. Information sources are often dependent, and this paper provides insight concerning the impact of dependence on the precision and value of information from a number of sources.

The results presented in this paper indicate that, in the context of the model studied, positive dependence among information sources can seriously reduce the precision and value of information. Except for some cases with extremely high correlations, independent information sources lead to greater precision and provide more valuable information than dependent sources. Furthermore, the differences between dependent and

independent sources can be surprisingly large. For example, if all of the sources have the same variance and the same pairwise correlation, an infinite number of sources with correlation 0.25 is equivalent to only four independent sources.

In terms of decision making, the expected value of information from independent sources is typically (but not always) a concave function of the number of sources. That is, each successive source of information results in a smaller increment in value of information than the previous source. With positive (but not extremely high) dependence, the incremental value of information decreases much more rapidly than in situations with independence. Moreover, no matter how many information sources are consulted, the value of information does not approach the expected value of perfect information.

What are the implications of these results for the decision maker who is contemplating the purchase of information from a number of sources? If the sources are moderately positively dependent, the rapid reductions in incremental gains in precision and value should lead the decision maker to purchase fewer sources than would be the case with less dependence. In fact, if the decision maker's own prior information is itself positively correlated with the information sources, then even a single source may not be a worthwhile purchase.

If additional information is desired in an inferential or decision-making problem, then it might be important to seek out information sources that are believed not to be highly correlated with each other or with the prior information. Even if this means using sources used that are less precise on an individual basis, trading some precision for reduced dependence can be advantageous. As we demonstrated, the marginal value of information decreases dramatically as the dependence increases.

ACKNOWLEDGMENT

This research was supported in part by the National Science Foundation under Grant IST8018578.

REFERENCES

- BATES, J. M., AND C. W. J. GRANGER. 1969. The Combination of Forecasts. *Opnl. Res. Quart.* **20**, 451-468.
- BLACKWELL, D. 1953. Equivalent Comparisons of Experiments. *Ann. Math. Statist.* **24**, 265-272.
- BUNN, D. W. 1978. *The Synthesis of Forecasting Models in Decision Analysis*. Birkhäuser Verlag, Basel, Switzerland.
- FIGLEWSKI, S. 1983. Optimal Price Forecasting Using Survey Data. *Rev. Econ. Statist.* **65**, 13-21.
- FIGLEWSKI, S., AND T. URICH. 1983. Optimal Aggregation of Money Supply

- Forecasts: Accuracy, Profitability and Market Efficiency. *J. Finance* **28**, 695-710.
- FRENCH, S. 1980. Updating a Belief in the Light of Some Else's Opinion. *J. Roy. Statist. Soc. Ser. A* **143**, 43-48.
- GEISSER, S. 1965. A Bayesian Approach for Combining Correlated Estimates. *J. Am. Statist. Assoc.* **60**, 602-607.
- GOKHALE, D. V., AND S. J. PRESS. 1982. Assessment of a Prior Distribution for the Correlation Coefficient in a Bivariate Normal Distribution. *J. Roy. Statist. Soc. Ser. A* **145**, 237-249.
- GRAYBILL, F. A. 1969. *Introduction to Matrices with Applications in Statistics*. Wadsworth, Belmont, Calif.
- HILTON, R. W. 1981. The Determinants of Information Value: Synthesizing Some General Results. *Mgmt. Sci.* **27**, 57-64.
- HOGARTH, R. M. 1977. Methods for Aggregating Opinions. In *Decision Making and Change in Human Affairs*, pp. 231-235, H. Jungermann and G. deZeeuw (eds.). Reidel, Dordrecht, Holland.
- LAVALLE, I. H. 1968. On Cash Equivalents and Information Evaluation in Decisions Under Uncertainty. *J. Am. Statist. Assoc.* **63**, 252-290.
- MORRIS, P. A. 1977. Combining Expert Judgments: A Bayesian Approach. *Mgmt. Sci.* **23**, 679-693.
- NEWBOLD, P., AND C. W. J. GRANGER. 1974. Experience with Forecasting Univariate Time Series and the Combination of Forecasts. *J. Roy. Statist. Soc. Ser. A* **137**, 131-165.
- PRESS, S. J. 1972. *Applied Multivariate Analysis*. Holt, Rinehart & Winston, New York.
- RAIFFA, H., AND R. SCHLAIFER. 1961. *Applied Statistical Decision Theory*. Graduate School of Business Administration, Harvard University, Boston.
- WINKLER, R. L. 1981. Combining Probability Distributions from Dependent Information Sources. *Mgmt. Sci.* **27**, 479-488.
- WINKLER, R. L., AND S. MAKRIDAKIS. 1983. The Combination of Forecasts. *J. Roy. Statist. Soc. Ser. A* **146**, 150-157.