Belgian Journal of Operations Research, Statistics and Computer Science, Vol 25, n° 1

COMBINING EXPERT OPINION UNDER A PSYCHOPHYSICAL LAW

Gerrit K. JANSSENS

Leerstoel Bedrijfsbeheer
Rijksuniversitair Centrum Antwerpen
Middelheimlaan 1, 2020 Antwerpen

ABSTRACT

A Bayesian approach to the combination of forecasts is reconsidered when experts do not necessarily produce unbiased estimates. The case of behavior under a psychophysical law is explored. An application in the area of medical decision-making is given.

1. Introduction

A decision-maker managing in a context of uncertainty often has to rely upon expert opinion. Consulting his pool of experts he is confronted to one of the following strategies:

- either rely upon the 'best' expert,
- either aggregate the opinion of different experts.

The problem formulation is identical to the one appearing in combining forecasts from different econometric of time-series models. It is generally known that combination of forecasts based on weighted averages has had considerable succes (see e.g. BATES & GRANGER (1969), DICKINSON (1973) and WINKLER & MAKRIDAKIS (1983)).

The reasoning behing the success of the second strategy is the fact that by choosing the first strategy the decision-maker is discarding independent evidence available from the rejected models or experts. From now on we will only proceed with the second strategy.

Consensus-type methods such as the Delphi-method (DALKEY, 1963) or a method by DEGROOT (1974) have been used to aggregate expert opinion. MORRIS (1974) claimed the need for a normative approach rather that to use those ad hoc methods. A Bayesian method first appeared in the time-series environment (BUNN, 1975), followed by its counterpart in the expert environment (MORRIS, 1977) and generalized by BORDLEY (1982a). In another article BORDLEY (1982b) proves that some of the ad hoc methods, in casu the BATES & GRANGER and DICKINSON formulae are special cases of the Bayesian approach.

The Bayesian as well as the ad hoc methods mostly assume that the estimates are unbiased, so that the sampling distribution is centered around the true value. But subjective judgment is a form of psychological behavior in which subjects seldom estimate unbiased. This bias can be described by the stimulus-response relationship known as Stevens' law:

$$\Psi = k \,.\, \Phi^{b}$$

stating that the perceived magnitude Ψ grows as the physical value Φ raised to the power b (for a review, see STEVENS, 1966).

In the second paragraph we will consider the consequences of the Bayesian combination of opinions of experts behaving under this psychophysical law.

In the third paragraph we will illustrate the theory with an application where medical subject groups evaluate the severity of different morbidity states.

2. The psychophysical law applied

Subjects giving an opinion on a socio-economic variable can also behave under the same stimulus-response relationship in spite of the fact that the physical stimulus is not very well defined: it has been approved by studies on the moral judgment of a number of offenses (EKMAN, 1962) and on judged frequencies (LICHTENSTEIN et al., 1978, and WARR, 1980).

Our theoretical model looks like:

$$z_k = a_k . x^{b_k} \qquad (1)$$

with z_k = individual (or group) k's assessment

x = the true value

a_k, b_k = individual (or group) k's parameters in Stevens' power law.

If a number of experiments m were realized, we could estimate both parameters by fitting the equation:

$$z_{ik} = e_i . a_k . x_i^{b_k} \quad (i=1, \ldots, m) \tag{2}$$

where the e_i, and by this also the z_{ik}, are lognormally distributed. With the requirement $E(e_i) = 1$, we see that the

$$\left[\frac{z_{ik}}{a_k} \right]^{1/b_k}$$

are lognormally distributed around the true value x_i. In the Bayesian approach BORDLEY (1982b) states that the combined estimate of expert assessments lognormally distributed around the true value is the geometric mean, assuming that in Bayes' formula the prior distribution may be omitted.

However, from a decision viewpoint a more interesting case is this in which the prior information is significant (or at least nog negligible) and the prior distribution in Bayes's formula cannot be eliminated (BERGER, 1980, p.131). BERGER (1980, p.93) presents a solution for the case in which expert estimates z_k are normally distributed around the true value x, and the prior distribution is N (x,v^2). In this case the posterior distribution can be written as:

$$f(x/z_1,\ldots,z_n) = \frac{1}{\sqrt{2 \pi\sigma^2}} . \exp \left[- 1/2 \left(\frac{x-Q^\circ}{\sigma} \right)^2 \right] \tag{3}$$

where $\sigma^2 = \left(\frac{1}{s^2} + \frac{1}{v^2} \right)^{-1}$

$$Q^\circ = \frac{1}{\sigma^2} \left(\frac{Q}{s^2} - \frac{x}{v^2} \right)$$

Q, s^2 are the group estimate and its variance as defined in BORDLEY (1982b).

To suit the conditions necessary to follow this approach, a transformation of the data is needed. Assuming the data allow the power law to be loglinearized, we obtain:

$$\log z_k = b_k \log x + \log a_k + \log e \tag{4}$$

The requirement $E(e) = 1$ applied to a lognormal distribution with parameters (α,β) leads to:

$$\exp\left(\frac{2\alpha + \beta^2}{2}\right) = 1 \tag{5}$$

and by this,

$$E(\log e) = -\beta^2/2$$

$$\sigma^2(\log e) = \beta^2$$

The ordinary least squares estimates are unbiased for b_k, but show a negative bias for a_k.

As there seems to be no satisfactory simple unbiased estimate for a_k, THOMAS (1981) proposed a two-step method:

(1) Build an improved estimate of $\log a_k$ - called $(\log a_k)_{I.E.}$ - by putting

$$(\log a_k)_{I.E.} = \widehat{\log a_k} + \hat{\beta}^2/2 \tag{6}$$

where $\log a_k$, $\hat{\beta}$ are the least squares estimates.
Intuitively, this improvement should result from the negative bias by an (expected) amount of $-\beta^2/2$.

(2) Use the improved estimate in the following way:

$$a_k = (1\text{-var}\left\{(\log a_k)_{I.E.}/2\right\}).\exp(\log a_k)_{I.E.})$$

THOMAS provides a replacement for the variance of $(\log a_k)_{I.E.}$ as it is not known. This two-step estimator for a_k is consistent, but negatively biased for small values of the number of experiments m.

If data can be gathered by means of paired comparisons the model can be reformulated in such a way that the estimation problem of a_k can be avoided. Since the tuples (a_k, b_k) are individual (or group) k's characteristics in responding to a stimulus, it is expected that a stimulus x' would result in an estimate z'_k given by

$$z'_k = a_k . x'^{\,b_k} \qquad (7)$$

Paired comparison allows us to write:

$$\frac{z_k}{z'_k} = \left[\frac{x}{x'}\right]^{b_k} \qquad (8)$$

in which case we only have to estimate b_k, for which the least squares estimator can be used. By raising the z_k/z'_k to the power $1/b_k$ we obtain unbiased estimates, which can be normalized, so that they can be used in the BERGER-BORDLEY formulae.

3. An application: judgments on the severity of morbidity states.

3.1. Problem situation

In an attempt to estimate hospital output, ROSSER & WATTS (1972) classified 29 morbidity states according to two dimensions, disability and distress. Subjects were interviewed to compare different morbidity states. The subjects were taken from 6 populations: medical patients, psychiatric patients, medical nurses, psychiatric nurses, healthy volunteers and doctors. In a paired comparison experiment subjects were asked to answer on questions as 'how many times more ill is a person described in state x as compared with state y'.

From these data ROSSER & KIND (1978) built a scale on $(0, +\infty)$, where 'fit' has value 0, 'dead' has a positive value and states worse than death are allowed.

To test the accuracy of the subject's views, in a later study KIND, ROSSER and WILLIAMS (1982) estimate scale values for the same morbidity states using data gathered from legal awards. The rationale for taking the legal scale as a basis for an objective estimate lies in the fact that the court has more time and more information to take its decisions.

3.2. The psychophysical law in the judgments

As an objective estimate is available, we can test the hypothesis that the subjects act in their judgment under the psychophysical law. Due to the fact that we use paired comparison-data, the slope of the power function has value 1 and has not be be estimated (TEGHTSOONIAN, 1973).

Results from least squares regression are shown in Table 1 and compared with explained variance in a linear model (these latter data from KIND, ROSSER and WILLIAMS, 1982).

TABLE 1

(1) Subject identification (k)	(2) Estimated coefficient b_k	(3) Standard error on b_k	(4) % variance explained in power model	(5) % variance explained in linear model
Medical patients	2.73	0.22	88	63
Psychiatric patients	3.13	0.28	86	63
Medical nurses	2.46	0.33	77	59
Psychiatric nurses	3.19	0.20	90	59
Healthy volunteers	3.73	0.09	98	64
Doctors	3.65	0.34	79	68

The fact that all coefficients are significantly greater than 1 does mean that the court assigned more balanced valuations to the more severe states. From a point of view of the explained variance, it can be seen that the power law in all cases is superior to the linear model.

The model includes the assumption that in the psychophysical model $a_k=1$ for all k. Through log-linearization the intercept term should be zero, but is mostly not, because it depends on the variance of the error term. By using a non-linear least squares algorithm (BARD, 1974), this bias could be excluded.

The data, previously fitted to x^{b_k}, are now fitted to ax^{b_k}. Results are presented in table 2.

TABLE 2

Subject identification	a_k	standard error on a_k	b_k	standard error on b_k
Medical patients	0.85	0.08	3.16	0.35
Psychiatric patients	1.12	0.10	2.85	0.36
Medical nurses	0.92	0.11	2.61	0.44
Psychiatric nurses	0.79	0.08	3.96	0.38
Healthy volunteers	1.04	0.05	3.61	0.17
Doctors	1.46	0.11	2.43	0.28

These figures show us that for the subject groups 'Psychiatric Nurses' and 'Doctors' the a_k - coefficient is significantly (95% confidence level) different from 1.

From a previous study (JANSSENS, 1985) we learn that agreement within the six subject groups is increased if one of those two subject groups is excluded. Agreement in this case was defined in terms of preference orderings in triads and measured by a coefficient of agreement proposed by KENDALL and BABINGTON-SMITH (see e.g. DAVID, 1969).

3.3. The procedure applied

In the following we only proceed with the 4 subject groups (Medical Patients= 1, Psychiatric Patients=2, Medical Nurses=3, Healthy Volunteers=4), whose behavior fits in the psychophysical law.

Using the coefficients b_k from Table 1, we transform the assessments z_k into unbiased assessments z_k^{1/b_k} and log-transform the latter to obtain a set of normally distributed values. From these we compute the variance-covariance matrix Σ^{-1} (row indices are the same as mentioned above):

$$\Sigma^{-1} = \begin{bmatrix} .27 & .34 & .33 & .22 \\ .34 & .42 & .41 & .28 \\ .33 & .41 & .38 & .27 \\ .22 & .28 & .27 & .19 \end{bmatrix}$$

The variance of the group estimate is:

$$s^2 = (I_N^T \; \Sigma^{-1} \; I_N)^{-1} = 0.20$$

where $I_N^T = (1,1, \ldots, 1)$.

As can be seen, the value of s^2 is nearly as small as the smallest of the individual variances (elements on the diagonal of Σ^{-1}).

The group estimate Q, following BORDLEY (1982b), is:

$$Q = (I_N^T \Sigma^{-1} \; I_N)^{-1} \; (I_N^T \Sigma^{-1} \; Z)$$

with in our case $Z=(z_1, \; z_2, \; z_3, \; z_4)$.

As the a posteriori distribution is normal with mean Q, Q can be considered as the maximum likelihood value and e^{Q} as the maximum likelihood value of the severity index.

An example is given:
The state belonging to 'Slight social disability' and 'Moderate Distress' (state (2,3)) has following scores z_k:

	kz_k	z_k^{1/b_k}	$\ln(z_k^{1/b_k})$
Medical patients	0.032	0.283	-1.261
Psychiatric patients	0.010	0.228	-1.479
Medical nurses	0.018	0.194	-1.641
Healthy Volunteers	0.027	0.378	-0.973
Combined estimate =			-1.375

This means a severity index of $e^{-1.375} = 0.253$.

REFERENCES

BARD, Y., 1974, Nonlinear parameter estimation, Academic Press, New York, 341 pp.

BATES, J.M. and C.W.J. GRANGER, 1969, The combination of forecasts, Operations Research Quarterly, vol. 20, pp. 451-468.

BERGER, J.O., 1980, Statistical decision theory, Springer Verlag, New York/Heidelberg/Berlin, 425 pp.

BORDLEY, R.F., 1982a, A multiplicative formula for aggregating probability assessments, Management Science, vol.28, pp.1137-1148.

BORDLEY, R.F., 1982b, The combination of forecasts : a Bayesian approach, Journal of the Operations Research Society, vol. 33, pp.171-174.

BUNN, D.W., 1975, A Bayesian approach to the linear combination of forecasts, Operations Research Quarterly, vol.26, pp.325-329.

DALKEY, N. and O. HELMER, 1963, An experimental application of the Delphi method to the use of experts, Management Science, vol.9, pp.458-467.

DAVID, H.A., 1969 (2nd ed.), The method of paired comparisons, Griffin & C°, London, 124 pp.

DEGROOT, M.H., 1974, Reaching a consensus, Journal of the American Statistical Association, vol. 69, pp.118-121.

DICKINSON, J.P., 1973, Some statistical results in the combination of forecasts, Operations Research Quarterly, vol.24, pp.253-260.

EKMAN, G., 1962, Measurement of moral judgment, Perceptive Motor Skills, vol. 15, pp.3-9.

JANSSENS, G.K., 1985, Agreement among judges on the severity of different morbidity states, RUCA Working Paper 85/04, Antwerpen.

KIND, P., R. ROSSER and A. WILLIAMS, 1982, Valuation of quality of life: some psychometric evidence, in M.W. JONES-LEE (ed.), The value of life and safety, North-Holland, Amsterdam, pp.159-170.

LICHTENSTEIN, S., P.SLOVIC, B. FISCHOFF, M. LAYMAN and B. COMBS, 1978, Judged frequency of lethal events, Journal of Experimental Psychology: human learning and memory, vol.4, pp.551-578.

MORRIS, P.A., 1974, Decision Analysis Expert Use, Management Science, vol. 20, pp.1233-1241.

MORRIS, P.A., 1977, Combining expert judgments: a Bayesian approach, Management Science, vol. 23, pp.679-693

ROSSER, R. and WATTS, V., 1972, The measurement of hospital output, International Journal of Epidemiology, vol.1, pp.361-368.

ROSSER, R. and KIND, P., 1978, A scale of valuations of States of Illness: is there a social consensus?, International Journal of Epidemiology, vol.7, pp.347-358.

STEVENS S.S., 1966, A metric for the social consensus, Science, vol.157, pp.530-541.

TEGHTSOONIAN, T., 1973, Range effects in psychophysical scaling and a revision of Stevens' law, American Journal of Psychology, vol.86, pp.3-27.

THOMAS, H., 1981, Estimation in the power law, Psychometrika, vol. 46, pp.29-34.

WARR, M., 1980, The accuracy of public perceptions about crime, Social Forces, vol.59, pp.456.

WINKLER, R.L. and SP. MAKRIDAKIS, 1983, The combination of forecasts, Journal of the Royal Statistical Society, Series A., vol. 146, pp.150-157.