DISCUSSION

P.J. HARRISON (*University of Warwick*):

In relation to Dr. Brown's paper, I would recall that this morning George Barnard mentioned the comparison done by I.C.I. Ltd. between Ridge Regression and ordinary Least Squares Regression. The recommendation arising from that study was to continue to use the latter.

Now one has to be very careful in making comparisons. Most of us will have suffered from what we would regard as totally unjustified comparisons. For example, consider the comparison of a sledge with a car as a means of transport. In the Artic the sledge with its husky dogs would tend to win whereas on a motorway in Spain the car clearly wins. But perhaps we are on the sea when neither a sledge nor a car is particularly useful! Consequently the major question about the current comparison concerns its relevance. Are we on ice, road or sea?

Multivariate analysis is always worrying since its successful application demands great care. In Phil Brown's example my worries are about robustness and model adequacy; about

(i) the global linearity with fixed coefficients over time;
(ii) associated Normality;
(iii) the structure of the variance matrix, particularly since the data are roughly proportions.

So by what standard can we judge the validity of the comparison? Examples give us a good opportunity for assessment and Phil is courageous enough to give us his data. Taking his specific election example, since one of the main purposes of election forecasting is to forecast as soon as possible the number of seats which will be won by each party, I looked at the 'Pred' comparison. Remembering that October 1974 and February 1974 are not far apart in time, I first postulated a 'no change' model M_0:-

'Each Party will retain a seat previously held'.

Looking at my table 1, this is seen to outperform all Phil's models as given in his table 3. Since M_o may be interpreted as a redundant

TABLE 1
A comparison of various estimators using Pred.
Showing a number of incorrect forecasts for varying n.

Number of results (n)	Least sq.	Mod H-K	Min Rid	Bayes S. Rid	M_o	M_1	M_2
0					5	5	5
15	9	11	7	6	4	4	1
25	7	9	6	5	4	4	1
45	4	3	2	2	3	2	1
65	0	0	0	1	0	1	1

ordinary Bayes regression model with an unshakeable prior, Phil's conclusions are clearly questionable if based on such examples.

We may investigate the comparison further using a very simple model in which we write for seat k

$$R_{ijk} = \theta_{ij} r_{ijk} + U_{ijk}$$

where party i won the seat in February and r_{ijk} is the number of votes then cast for party j relative to those cast for party i. R_{ijk} is the corresponding quantity for October. θ_{ij} is an unknown regression coefficient and U_{ijk} a Normal random variable with zero mean and here the variance is inappropriately taken as a constant V.

Thus the model can be written over all the constituencies as

$$\mathbf{R} = \mathbf{r}\theta + \mathbf{U} \qquad\qquad \mathbf{U} \sim N(\mathbf{0}; \text{diag}\,(\mathrm{V}))$$

where $\mathbf{R}$ and $\mathbf{V}$ are the vectors of all the meaningful R_{ijk}'s and U_{ijk}'s, $\mathbf{r}$ is a matrix with only one appropriate non zero quantity in each row and θ is the column vector of the θ_{ij}'s. For model M_1 we will take an exchangeable ignorance prior structure at time $t=0$ as

$$(\theta\,|\,\mathrm{t}=0) \sim N\,[\mathbf{1};\ \text{diag}\ 10^{100}]$$

Thus M_1 effectively performs independent least squares regressions in sequentially estimating each of the elements of θ. At any time the Pred forecast for a winner of a constituency seat is that party with the highest expected proportion of votes. The performance of M_1 is given in Table 1.

The particular purpose of this election was for the Labour Party, who held a small majority of seats, to go to the country and obtain an increased majority. Consequently rather than the ignorance prior on θ of model M_1 it could be argued that although such a prior would be suitable for seats held by parties other than the Labour Party, there was a strong priori argument to say that Labour would retain seats it previously held. If this is so then perhaps a more realistic prior for θ would have been

$$(\theta\,|\,\mathrm{t}=0) \sim N\left[\begin{pmatrix}\mathbf{1}\\ \mathbf{1}\end{pmatrix};\quad \begin{pmatrix}\text{diag}\,10^{100} & \mathbf{0}\\ \mathbf{0} & \text{diag}\,\epsilon\end{pmatrix}\right]$$

where the lower block only relates to seats held in February by Labour and where ϵ is small. With such a model M_2, the performance on Pred is extremely good as shown in the table and again outperforms all others with a great deal to spare.

Whether or not one accepts model M_2, the performance of the elementary models M_0 and M_1, and the results of the I.C.I. study seriously question the authors conclusions and stress the need for great care in making comparisons of technique particularly when one of them appears to lack robustness to variation in its assumptions.

There has been little time to study Professor Dempsters paper and I will just consider a few points. He asks is there such a thing as a good or valid method of seasonal adjustment? I suppose that in order to answer this we need to be very clear what we mean by seasonality and thus I would ask what does it mean?

In his case, at the Bureau of Census, if a deseasonalised series of official statistics is to be produced for many anonymous users with a variety of utilities, some of them are going to want to know what has been filtered out of the data and what has not.

For example in my own implementation of short-term forecasting situations in which we use structured state-space or Dynamic Linear Models we discuss with users what will and what will not be parameterised and what it will mean. This is particularly important with respect to seasonality. If we are considering monthly temperatures in Central England (Box-Jenkins) then we might be happy to describe the seasonal effect by a first harmonic describing the effect of the elliptic orbit of the earth round the sun. But what about 1976 when there was a very hot summer and a drought. In order to get a deseasonalised figure do we take out the 'average seasonal factor over the years or do we somehow speculate that we have had a particularly hot summer but that this is no reason to think that the so called deseasonalised level has changed? To me you take your choice dependent upon your use of the resulting figures. But of course in Arthur's case he does not know all the uses to which the figures will be put. To be more concrete on this point, one of our recent clients who is a manufacturer of alcoholic drink has experienced a major rise in demand over 1975 and 1976. This was naturally attributed to great marketing success although what had really happened largely resulted from the unusually hot couple of months in the two years combined with an advantage relative to the other alcohols of no V.A.T. When the summer weather reverted to its more typical pattern and the brewers successfully lobbied for V.A.T. on this particular drink, sales fell dramatically. Clearly in this sort of case it is of vital importance to the Company to estimate how much of its sales is to be thought of as due to relative price advantage, promotions, variation within year etc. But again what is the deseasonalised series? Do we take out the seasonality of advertising, of the earth going round the sun, of the particular weather conditions giving rise to a freak summer, the seasonal buying habits of the customer, which in the case of government institutions may just reflect the current policy of placing orders regularly at quarterly intervals and not any seasonal usage, and so on.

If asked why they like deseasonalised series, many users would remark that they are looking for turning points and trend changes. Perhaps also they are now interested in other forms of change such as jumps in 'underlying level'. For most official series it is probably good enough to filter out regular seasonal effects in the traditional ways, using local linearity and smoothing. However, this is not adequate to deal with sharp changes in the series or in any of its structural components. The absence of any parametric formulation in Professor Dempsters paper surprises me. At this meeting we have heard from Professor Akaike, Adrian Smith and Dr. Makov about such formulations and of course Colin Steven and I have used these successfully for many years under the title of Bayesian Forecasting (1976). The big advantage of the parametric or state-space formulation is that it allows one very easily, to attribute variation to many sources and is, after all, in the spirit of the Bayesian statistical linear model with the associated estimation of effects in terms of their distribution functions.

We have also had success with this Bayesian approach in hierarchical forecasting and gave a paper on this at the Royal Statistical Societies Conference in 1977 (Harrison, Leonard and Gazzard). Here we were commissioned to develop a method of forecasting aggregates and their constituents ensuring compatibility in the sense that the sum of the parts was equal to the whole at all levels. Again with probability distributions this reduces to conditional forecasting and this method has been fully documented by I.C.I. who use the resultant software a great deal.

However in expressing my surprise at the ommission of this type of representation let me admit that I am unfamiliar with Mandelbrot's approach and that since I only got a copy of Arthur's paper yesterday I have had no chance to look up the literature. At first glance I am not attracted to it and on a technicality with equation 3.3 would ask why with $H = 0.5$, the variance $C(0) = O$?

I would close with one important point. It is my personal view that stationarity imprisons us. Surely we must recognise that variables have probabalistic effects. That is if I launch an advertising campaign described only in terms of the amount of money to be spent I am unsure of its effect and hence my view of the future is much more uncertain. I can learn on line about this probabalistic effect and I can model it or I can pretend a naive stationarity and ignore it. The most convenient way of modelling that I know is the parametric approach and I would urge that this be given much more attention.

A. ZELLNER (*University of Chicago*):

The papers by Brown, "Aspects of Multivariate Regression" and by Dempster, "Bayesian Inference in Applied Statistics" are valuable contributions which treat important problems. Since I prefer reverse alphabetical order for an obvious reason, I shall comment on Dempster's paper first and then turn my attention to Brown's paper.

First, Dempster presents a personal view of how Bayesian inference ought to be implemented. According to Dempster, "The objective of Bayesian inference is to quantify uncertain knowledge about a set of unknown quantities in terms of a posterior distribution of those unknowns." (p.1). In connection with this definition, it is important to include as yet unobserved values of variables in the "set of unknown quantities" and to mention predictive distributions. In addition, I believe that Dempster's discussion would be enriched by relating it to various theories of scientific method, for example Jeffreys's and those of other philosophers of science. Without such considerations, it is difficult, if not impossible, to appraise Bayesian and other systems of statistical inference. For example, the issue of whether probability is better regarded as a frequency or non-frequency concept requires analysis in terms of a theory or alternative theories of scientific method.

As regards some specific issues in Dempster's discussion of Bayesian inference, his remarks on "tail area significance tests" should, in my opinion, be expanded to consider Jeffreys's Bayesian significance test procedures. As pointed out in my and Siow's paper for this Conference, for many testing problems Jeffreys-like posterior odds ratios relating to pairs of hypotheses are monotonically increasing functions of tail areas associated with usual t and F statistics used by non-Bayesians in testing hypotheses. Thus there is a direct link between Jeffreys-like posterior odds ratios and tail areas, a relationship which may explain why many applied statisticians have persisted in their use of tail areas or "p-values" in appraising hypotheses.

I am sympathetic to Dempster's view that "Bayesian inference is logically separable from decision analysis". De Finneti at a 1968 conference at Frascati expressed a similar view and I believe that R.A. Fisher and H. Jeffreys also agree with

Dempster's view. However, I.J. Good and others appear to take the position that "quasi-utilities" are generally employed in inference and that inference may be viewed as partially contained within decision analysis. Clearly there is a need for more work on the axiom systems underlying inference (or learning) and decision (or utility) analysis to help resolve the issue of separability. For example, the relation of Jeffreys's and Savage's axiom systems could be studied to determine whether the separability view is logically tenable.

Second, in connection with Dempster's discussion of the practical problem of modelling seasonal time series, let me draw participants's attention to the recently published volume, Zellner, ed. (1978), in which many of the issues discussed by Dempster are treated at length in contributions by a number of statisticians and econometricians. In the volume, three approaches to the analysis of seasonal time series are distinguished, namely (1) the descriptive, non-modeling approach, (2) the statistical modeling approach and (3) the subject-matter causal modeling approach. Dempster's approach is an ingenious example of a statistical modeling approach that he compares with two other statistical modeling approaches, the stochastic seasonal ARIMA approach associated with Box and Jenkins and an AR approach advocated by Parzen and Hiplel and McLeod. One might add to this list the mixed deterministic-stochastic seasonal models developed and applied by Pierce, (1978). While these statistical modeling approaches can yield useful results, it is my view that they must be augmented by a subject-matter causal modeling approach. Without a good subject-matter understanding of the nature of seasonality and factors which produce changes in seasonal patterns, it is the case that mechanistic, statistical models do not have a firm foundation.

Practically speaking, this means that parameters of such models may in fact be variables and thus the difficult problem of assessing good prior distributions for these "parameters" may be intractable. Further, as Ploser (1978) notes, analysis of economic models makes it highly unlikely that the restrictions on the moving-average polynomial-lag operators needed to produce the Box-Jenkins multiplicative seasonal ARIMA schemes will be satisfied in general. Also, the variation of policy-control variables can introduce non-stationary effects in the time series processes for individual variables. These considerations, and others which could be added, point in the direction of devoting more effort toward understanding the causes of seasonality and producing reasonable, serious subject-matter models which will enhance our scientific understanding of seasonal phenomena, for example changing seasonal patterns and differences in seasonal processes for different variables. Bayesian techniques can be employed to appraise alternative subject-matter models, estimate their parameters, and use them for prediction and policy purposes. It could very well be the case that fractional Gaussian processes will be valuable in the context of subject-matter causal modeling of seasonal time series. My impression is that Dempster appreciates these points and plans to devote more attention to them in future research.

I am in full agreement with Dempster on the importance of computation in Bayesian analyses and the need for good Bayesian computer programs. In connection with our NBER-NSF Seminar on Bayesian Inference, we have established a Computation Committee headed by Joseph B. Kadane. S. James Press, a member of the Computation Committee has written a paper, Press (1980), which provides information about a number of computer programs.

With respect to Brown's paper, he is critical of the usual least squares estimate or diffuse-prior posterior mean for the regression coefficients shown in (2.3) and (2.4) of his paper on grounds that it does not take account of the between regressions covariance matrix, $\Gamma = \{\gamma_{ij}\}$. This is not a reasonable critique since (2.3.) is the posterior mean relative to a particular prior and the maximum likelihood estimate based on the normal and other symmetric distributions for the error terms. The fact that the symmetry of the problem results in the regression coefficient estimates not depending on the nuisance parameters in Γ seems to be a blessing and not a fault. Also, Hill's cogent discussion of near admissibility or restricted admissibility of "usual" estimates, such as $\hat{B} = (\hat{\beta}_1, \hat{\beta}_2, \ldots, \hat{\beta}_q)$ in (2.3), indicates circumstances in which $\hat{B}$ can be justified as an estimate —see Hill's paper, cited by Brown, pp. 566-568—. Whether these particular circumstances obtain is the crucial issue. If they do not, then it is unreasonable to use $\hat{B}$; if they do, it is reasonable to use $\hat{B}$. As Hill remarks, "Thus stable estimation may justify the use of Lebesgue measure and the estimator Y [here, $\hat{B}$], as approximations... but it is important to be aware that the approximation must be justified separately in each usage and that it cannot hold for all y." (p.568). Hill talks of "approximations" because he believes that some prior information is available. On the other hand, Jeffreys views Lebesgue measure as a canonical prior for representing ignorance regarding the values of the regression coefficients and $\hat{B}$ the appropriate, not approximate, estimate given the assumed state of ignorance. Of course, if more information is available, as assumed by Hill and by Stein, it can be employed and will lead to estimates of $B = (\beta_1, \beta_2, \ldots, \beta_q)$ which are not independent of Γ, assuming, as Brown does, that Γ has a known value. For example, in the case of exact linear restrictions on the elements of B, it is well-known that the maximum likelihood estimate will usually depend on Γ.

Rather than assess a serious prior for the elements of B, a chore which Brown considers too onerous, he opts for use of the exchangeability assumptions described on p. 5. It is important to emphasize that these assumptions are hard to defend in many practical applications. In the Lindley-Smith approach $\beta_i \sim N(\mu, \Sigma)$ implies that all q regression coefficient vectors have the same mean μ, hardly a satisfactory assumption in many economic applications. Further, the assumption that $\mu_i \sim N(\gamma, \sigma_\gamma^2)$ implies that the elements of μ have a common mean, again hardly a tenable assumption in many applications. Also, the zero mean assumption in (3.3) leading to (3.4) is tenuous in many applications. However, Brown notes that the zero mean assumption can be relaxed and also writes, "Of course the appropriateness of priors depends on the application..." (p.6). Thus Brown emphasizes, quite reasonably, that one must assess the appropriateness of prior assumptions and I contend this process is not far different from assessing an appropriate prior distribution for the regression coefficients.

I pointed out some years ago that the restrictiveness of the natural conjugate prior for the elements of B, mentioned in Rothenberg (1963) can be avoided by assessing a general normal prior for the elements of B —see Zellner (1971), pp. 238-240.— If we write $\beta' = (\beta_1', \beta_2', \ldots, \beta_q')$, the prior which I suggested is:

$$p(\beta,\Gamma) \propto |\Gamma|^{-(q+1)/2} \exp\{-(\beta-\bar{\beta})' C^{-1} (\beta-\bar{\beta})/2\} \tag{1}$$

a diffuse prior for the elements of Γ (or Σ in my notation), and a normal prior for the pq elements of β, with prior mean $\bar{\beta}$ and prior covariance matrix C. Using this prior, I derived the following approximate posterior mean, **b**, for β:

$$\begin{aligned} \mathbf{b} &= (C^{-1} + S^{-1} + X'X)^{-1}[C^{-1}\bar{\beta} + (S^{-1} + X'X)\hat{\beta}] \\ &= \bar{\beta} + [I_{pq} - (C^{-1} + S^{-1} + X'X)^{-1}C^{-1}]\,(\hat{\beta}-\bar{\beta}) \end{aligned} \tag{2}$$

where $\hat{\beta}' = (\hat{\beta}_1', \hat{\beta}_2', \ldots, \hat{\beta}_q')$ and $S = (Y-X\hat{B})'(Y-X\hat{B})/n$.

It is seen that **b** is a matrix-weighted average of the prior mean vector, $\bar{\beta}$, and of the least-squares estimate, $\hat{\beta}$, with their respective precision matrices as weights. The second line of (2) puts **b** in a "shrinkage" form where the shrinkage is toward the prior mean vector $\bar{\beta}$. If it is appropriate to specialize (2), for example by giving C a particular form or by assuming $\bar{\beta} = \underset{\sim}{0}$, it is possible to obtain particular "ridge-like" estimates. The critical issue is whether these particular specializing assumptions are reasonable. If they are not, it is unreasonable to impose them. Also, it should be mentioned that (2) is the mean of an approximate normal posterior distribution for β with covariance matrix $(C^{-1} + S^{-1} + X'X)^{-1}$. With additional effort, a better approximate posterior distribution for β could be obtained.

Two issues arise regarding the prior in (1). First, it would be useful to have an informative prior for the elements of Γ. Ando and Kaufman pointed out that a prior in the inverted Wishart from places strong restrictions on the prior variances and covariances of the elements of Γ. While Lindley and Press have made some progress on the problem of formulating an informative prior for Γ, I do not believe that the problem has been satisfactorily solved. As regards procedures for assessing the normal prior for β in (1), an extension of the approach (Zellner, 1972) which I formulated for assessing normal priors in univariate, multiple regression models is possible. Applying this approach to each regression equation yields the prior mean and covariance matrix for each β_i, $i = 1, 2, \ldots, q$, namely $\bar{\beta}_i$ and C_{ii}, the matrices on the diagonal of $C = \{C_{ij}\}$. The extension of the approach involves the assessment of C_{ij} for $i \neq j$, that is $\mathrm{cov}(\beta_i, \beta_j)$, $i \neq j$*. Brown's paper has stimulated me to consider this problem which I regard as tractable.

In summary, I urge Brown and others who utilize procedures based on exchangeability assumptions or ridge-regression procedures to consider carefully the assumptions underlying their procedures. I believe that careful attention to these assumptions will lead to a serious assessment of prior distributions which is required to avoid introducing erroneous information in analyses.

* Further, the assessment procedure can and should include checks on the assumed normal form of the prior.

REPLY TO THE DISCUSSION

P.J. BROWN *(Imperial College, London):*

I am most grateful for the two invited discussions presented here and the verbal contributions from various participants at the symposium. Professor Zellner has a number of points concerning the first part of my paper. Let me comment generally on these. The motivating force of our work is the prior distribution for the regression coefficients in multivariate regression. It seems important to me to delineate sets of archetypal priors, investigate their implications, choosing between these priors, in a practical situation by means of my prior knowledge for the particular situation together with accumulated knowledge of the implications of divergence of behaviour should the prior be inappropriate. Indeed L.J. Savage (unpublished book, The Subjective Basis of Statistical Practice, 1961, Section 2.15) emphasises the fuzziness of held prior opinions. He states "In practical work, I try to take advantage of whatever common properties of the acceptable probabilities I can discern". Exchangeability is a very important feature, valid in some situations but not in others as emphasised in the paper. Furthermore, results such as the sampling theory results of section 4 enable one to investigate theoretically the performance of a class of estimators resulting, relative to the Bayes estimator which corresponds to vague prior knowledge. Synthesis of Bayesian and sampling theory properties is, I think, important.

I very much appreciate the substantial contribution from Professor Harrison. He has indeed hit the nail on the head in questioning the general relevance of least squares and his models deserve careful consideration. I naturally disagree with the nature of his criticism of 'ridge regression'. I think 'ridge regression' provides a range of possibilities of wide but of course no means universal usefulness. Let me answer some of his points in detail.

I echo Jeff's concern for careful comparison, particularly in the I.C.I. study he mentions. I do not wish to criticise the company that gave us both sustenance for a number of years; rather the summary nature of the conclusions stated here. Both 'ridge regression' and even ordinary least squares are not simple well defined techniques. Their application to a practical problem involves various protocols such as, which variables to include, whether to transform them, etc. Additionally 'ridge regression' demands careful scaling of the explanatory variables guided by prior information.

Also of crucial importance is the estimation of the ridge constant. Many of the methods of estimating this constant mentioned in our paper were not available when the I.C.I. study was concluded in the early 1970's. In the absence of detailed evidence I must therefore be rather sceptical about the study.

As far as the study in our paper is concerned one does need to beware of using the PRED criterion in isolation. Continuing Jeff's nautical metaphor, it is a bit of a red herring. Its virtue of simplicity of calculation masks the fact that it is really the probabilities of winning each seat that is important. These probabilities may be summed to give overall predictions. Integrations necessary for their calculation are performed in election night forecasting for the B.B.C. (Brown and Payne, 1975) but I did not go to the trouble of calculating them in this paper. In their absence goodness of prediction is better reflected by such measures as *SD* which measure the closeness of observed and predicted values. For this small subset of the full 635 constituencies in the

United Kingdom there are many models that do well retrospectively using the criterion PRED. In fact a very simple well established model gives the same performance on PRED as M_1, the most flexible of Jeff's models. This model estimates the percentage change for each party and adds the average of these changes uniformly to all the undeclared constituencies. The calculations for $n = 15$ observations declared are given in Table 4 in the fine detail necessary to assure oneself about what is happening in the data.

TABLE 4

Votes (in thousands) and Percentage Changes of Electorate for each party from February to October 1974:

CONSTITUENCY	Electorate (thousands)	C2	C1	% ΔC	S2	S1	% ΔS	L2	L1	% ΔL	N2	N1	% ΔN
KILMNOCK	60	9	14	-8	22	24	-3	3	5	-3	15	8	12
GL CEN	26	2	3	-4	9	9	0	1	1	0	3	2	4
FIFE E.	56	16	21	-9	7	7	0	5	8	-5	13	9	7
DUMFRIES	62	18	22	-7	13	13	0	4	6	-3	13	9	7
G.L. PROVN	55	3	6	-6	21	23	-4	*	*	*.	11	7	7
G.L. SHETL	38	4	6	-5	13	14	-3	1	*	*	7	6	3
G.L. SPRING	48	4	7	-6	17	18	-2	1	*	*	9	8	2
G.L. CATH	50	16	18	-4	15	16	-2	1	*	*	6	5	2
EDBR N.	47	13	16	-6	8	9	-2	4	5	-2	8	5	6
G.L. GARSC	55	5	10	-9	20	21	-2	2	*	*	12	9	6
AYR	52	17	22	-10	14	17	-6	3	*	*	7	5	4
EDBR PNT	55	14	18	-7	13	14	-2	4	7	-6	10	5	9
GL KELVIN	43	7	11	-9	12	13	-2	2	*	*	6	6	0
GL MARYHL	52	3	7	-8	20	20	0	1	*	*	10	9	2
ANGUS S	52	15	21	-12	4	6	-2	3	*	*	17	15	4
	Averages			-7			-2			-3			+5

Key: * Denotes to contending candidate for that party

Notice that the percentage changes in proportion of the electorate are reasonably stable. Conservative votes are tending to go down by about seven percent, Labour by about two percent, Liberal by about three percent and Nationalists go up by around five percent. There is some evidence that Glasgow constituencies (listed as GL) move less towards the Nationalists (only a three percent increase).

With these estimated changes four constituencies are wrongly predicted on PRED.

GL GOVAN and STIRL FG are wrongly predicted as changing hands from Labour to Nationalist. Perth and EP is wrongly predicted as remaining Conservative and finally ROSS and CRM is wrongly predicted as switching from Conservative to Nationalist. Also, aside from re-estimation of average changes as n increases, since all but the last of these four constituencies fall in the 25 to 45 declaration band they will not cause prediction problems after $n = 45$.

Noting that the constant term is generally left unshrunk (section 5.1) this model is actually our ridge model with a very large ridge constant. With a lower value of ridge constant allowance is made for possible dependence on the other variables and indeed the variable for the incumbant party in February is typically of some usefulness in prediction. Looking back we might have envisaged that a Labour incumbancy would have been more valuable than a Conservative incumbancy but the effect is minimal on criteria other than PRED. Overall let me say however that in the election night forecasting context of predicting 635 constituencies, prior information from a multitude of sources is used and more variables entertained than used here. Of some importance, for example, are variables which define the perceived tactical situation in a constituency, for if ones favourite party stands little chance in your constituency you might decide to vote against the party you don't like by voting for the party that you dislike less. In general local information is available from psephological experts, opinion polls and 'post' polls. For details of Election night forecasting as implemented for the B.B.C. see Brown and Payne (1975). The example in this paper does not claim to reflect the multitude of concerns to be found there. Additional experience in the two 1979 elections (General and Direct Election to European Parliament) will be reported shortly.

A valid point to come out for our present paper is I think that the method of estimation of the ridge constant is critical and notwithstanding the success of our study, when the number of observation is small compared with the number of parameters, methods such as HKB and Sclove do mimic least squares too closely. They are not applying the implied prior information strongly enough in these circumstances. In anticipation, in all election work for the B.B.C. we have used a fixed ridge constant specified from previous experience. Reiterating, ridge regression is not a single universal tool but requires careful molding to available prior information.

Our experience of election forecasting does give us confidence that the concerns of 'robustness' and model adequacy as listed under (i), (ii) and (iii) of Jeff's discussion are not compelling. Indeed it seems that here Jeff himself does not find them compelling since his models M_1, M_2, involve the additive assumption of normal homoscedastic error and constancy of parameters over time. Multiplicative analysis of proportions in the voting context is fairly well established. Hawkes (1969) considers various models accounting for the transitions between parties from one election to another. These models are typically rather unstable with respect to election data and seem not to be much used although Miller (1972) has used a version of ridge regression to aid estimation. Model M_1 is somewhat simpler than those models in that in much the same spirit as the ubiquitous 'swing' (average of party i increase and party j decrease in the share of the total vote (or two party vote) it concentrates on the ij transition without accounting for the effects of other parties.

Both models M_1, M_2 have the air of prior distributions constructed retrospectively so as to perform well on PRED. Both models concentrate on the previous winner (party i) and provide no linkage between θ_{ij} for different i. Thus after fifteen constituencies have declared, all of which had been Conservative or Labour held, no information is available on θ_{ij}, i denoting Liberal or Nationalist. Total reliance is on the prior.

If there is a case for a change of model it is that votes be assumed to be Poisson (before conditioning on the electorate size) and that the log of the votes be linear in a set of variables. The computational burden of such a log-linear approach does not seem to be very necessary over the typical range of election data as evidenced by even the simple calculations of Table 4. Also Model M_1 with additive error is to me a little unnatural. Modified to a multiplicative error all the calculations of our present paper could be applied, if so desired, to changes in log proportion of electorate. Further, if S_{jk} is the ratio of October to February votes for party j in constituency k, the model

$$\log S_{jk} = \beta_{Jk} + e_k$$

where β_{jk} has a linear structure in terms of explanatory variables (which could include i, the previous winning party) naturally leads to an appealing measure of ''log-swing' from party j to party 1 given by

$$\beta_{jk} - \beta_{1k}.$$

However, it is easy to lose the inclination to study such modifications when our existing prediction methods perform as well as evidenced in the recent Direct Elections to the European Parliament. I hope you did not miss the B.B.C. programme 'Decision for Europe' (June 10th 1979) which presented them.

A.P. DEMPSTER (*Harvard University*):

I thank Professors Harrison and Zellner for their wise comments, most of which I take as friendly amendments.

I have, over 25 years, spent much time studying the views of many past and present leading thinkers on inference, feeling close to some such as Fisher and deFinetti and more distant from others such as Jeffreys and Savage. I hope one day to develop a reasoned exposition of my position, including its almost total debt to others. For now, however, I think more is to be gained by using the actuality of experience in applied statistics to inform theories of scientific methods than vice versa. In particular, the use of axioms to buttress a largely transparent logical system seems to me less valuable than extended testing of the consequences of the axioms in practice.

Zellner wishes that I would use more econometric theory and causal modelling, and I certainly hope to do so. I do wonder, however, whether the so-called causal models of macro-econometrics have much to do with the real causal factors which necessarily operate at a very micro level. The problems of inadequate information to specify a realistic causal system are so great that causal interpretations of feasible macromodels may do more harm than good, if taken at all seriously.

I agree with Harrison that seasonal techniques should be documented publicly.

They also should be defended rationally, which I think means having their Bayesian origins exposed. A good technique needs to make a rational assumption about how much one hot summer should affect one's judgments about the following summer. If good climatic theories (e.g., about ocean temperatures and currents) are available, they should be used, but in the end there will be a residual dependence on unverifiable prior assumptions. Ideally, several different scenarios should be presented so that the naive user can be warned and the sophisticated user can introduce his own prior beliefs.

REFERENCES IN THE DISCUSSION

BOX, G.E.P. and JENKINS, G.M. (1976). *Time-Series Analysis: Forecasting and Control.* San Francisco: Holden-Day.

HAWKES, A.G. (1969). An approach to the analysis of electoral swing. *J. Roy. Statist. Soc. A* **132,** 68-79.

HARRISON, P.J. and STEVENS, C.F. (1976). Bayesian Forecasting (with discussion). *J. Roy. Statist. Soc. B* **38,** 205-247.

HARRISON, P.J., LEONARD, T. and GAZZARD (1977). Multivariate hierarchical forecasting. *Proc. Annual Conf. Roy. Statist. Soc.,* Manchester 1977.

MILLER, W.L. (1972). Measures of electoral change using aggregate data. *J. Roy. Statist. A* **135,** 122-142.

PIERCE, D.A. (1978). Seasonal Adjustment When Both Deterministic and Stochastic Seasonality are Present. In *Seasonal Analysis of Economic Time Series,* (Zellner ed.) 242-269, Washington, D.C.: U.S. Government Printing Office.

PLOSSER, C.I. (1978). A Time Series Analysis of Seasonality in Econometric Models. In *Seasonal Analysis of Economic Time Series,* (Zellner ed.) 365-397. Washington, D.C.: U.S. Government Printing Office.

PRESS, S.J. (1980). Bayesian Computer Programs. In *Bayesian Analysis in Econometrics and Statistics: Essays in Honor of Harold Jeffreys,* (Zellner ed.) 429-442 , Amsterdam: North Holland Publishing Company.

ROTHENBERG, T. (1973). A Bayesian Analysis of Simultaneous Equation Systems. *Tech. Report.* **6315,** Rotterdam: Econometric Institute.

ZELLNER, A. (1971). *An Introduction to Bayesian Inference in Econometrics.* New York: Wiley.

— (1972). On assessing informative prior distributions for regression coefficients. *Tech. Rep.,* University of Chicago.

— (1978). (ed.) *Seasonal Analysis of Economic Time Series.* Washington, D.C.: U.S. Government Printing Office.