A Modified Inflation Procedure^{1,2}

EDWARD S. EPSTEIN

The University of Michigan, Ann Arbor

23 September 1966

In a recent note Glahn and Allen (1966) pointed out that "inflation" procedures of regression forecasts, as suggested by Klein *et al.* (1959, 1965), and others, necessarily have some undesirable attributes. Most significant, the inflation procedure necessarily increases the error variance of the forecast by introducing a bias which increases systematically as the forecast value of the predictand deviates further from its mean. However, the procedure does have the desirable (in most cases, but not in all) attribute of producing forecast values of the predictand which will more frequently take on extreme values. The standard deviations of the predicted and observed values of the predictand are made equal by this method, at least in the dependent sample, whereas the standard deviation of the uninflated predictand is always smaller.

The reason for employing a technique such as this is a prior conviction that the usefulness of a forecast depends more on the ability to "hit" the extreme or unusual events than it does on accuracy of the forecast *per se*. Expressed in another way, the use of this inflation procedure implies that one is more willing to accept relatively large errors which usually are in the direction of "over forecasting," than smaller errors of unbiased direction. In standard regression procedures, errors toward or away from the mean are equally likely. As Glahn and Allen point out, this judgment can be tested only in terms of users' utility matrices. Since such matrices are generally not known, however, standard techniques based on our best judgment concerning their nature must be substituted for the more specific techniques which could be developed in terms of such matrices.

It is my judgment that the considerations which led to the introduction of the inflation techniques are valid. Errors directed away from the mean tend to be less costly than those directed toward the mean. I would therefore suggest a regression method which recognizes this, but nevertheless is free from the inflation procedure's built in biases. The method involves the introduction of a scaled dependent variable, where the scaling is non-linear and varies in such a way as to allow differences in its value to be more nearly proportional to the

FIG. 1. Scaled vs. unscaled predictand.

"seriousness" of an error. If y_s is the scaled version of y , then one would expect $d^2y_s/dy^2 \leq 0$ when $y \geq \bar{y}$, in order to express the considerations mentioned above as to the seriousness of the error. One such scaling function which has this attribute is

$$y_s = F s_y \operatorname{sgn}(y - \bar{y}) \ln \left\{ \frac{|y - \bar{y}|}{F s_y} + 1 \right\}, \quad (1)$$

where

$$\operatorname{sgn}(z) = \begin{cases} +1 & z > 0 \\ 0 & z = 0 \\ -1 & z < 0 \end{cases}$$

s_y and $\bar{y}$ are the standard deviation and mean of the predictand in the dependent sample, and F is an arbitrary parameter which controls the amount of curvature in the relationship between the scaled and unscaled predictand. Eq. (1) is plotted in Fig. 1 for several values of F . The function is continuous and has a continuous first derivative, but its second derivative is discontinuous at the origin.

Having introduced the scaled dependent variable, one proceeds in the usual way to develop a regression equation in terms of the independent variables, x_i , i.e.,

$$\hat{y}_s = a_0 + \sum a_i x_i,$$

where the a_i are the usual least squares regression coefficients, and finally

$$\hat{y} = \bar{y} + F s_y \operatorname{sgn}(\hat{y}_s) \{ \exp[|\hat{y}_s|/F s_y] + 1 \}$$

is the estimate of the dependent variable.

¹ This research was supported in part by the U. S. Department of Commerce, Environmental Science Services Administration, Weather Bureau, under Contract Cwb-10847.

² Publication Number 111 from the Department of Meteorology and Oceanography, The University of Michigan, Ann Arbor.

FIG. 2. Computed regression lines (normal, inflated, and scaled) for values of $F = 0.1, 0.55$ and 1.0 . Each set of lines is based on 1000 "observations" from a bivariate normal distribution with $\rho = 0.5$.

FIG. 3. Computed regression lines (normal, inflated, and scaled) for values of $F = 0.1, 0.55$ and 1.0 . Each set of lines is based on 1000 "observations" from a bivariate normal distribution with $\rho = 0.8$.

The resulting estimates have certain advantages and disadvantages in comparison with those derived by other procedures. One no longer is postulating a linear relationship between the dependent and independent variables. If a linear relationship is valid, then the scaling procedure will introduce errors of estimation which will tend to be most serious about one standard deviation from the mean and for very extreme values. On the other hand there is in general no *a priori* reason to expect the dependent variable to have a linear relationship to the independent variables, any more than one may expect the scaled version of the dependent variable to bear a linear relationship to the predictors. If the latter is the case, i.e., if the relation to the scaled predictand is more nearly linear, the scaling procedure may actually result in a smaller mean square error than the normal regression method. As compared to the inflation method, the scaling procedure is likely to produce smaller errors of estimation except when the multiple correlation coefficient is very near one. If one assumes the scaled dependent variable is indeed linear in the independent variables, then the errors of estimate of y are equally likely to be positive or negative, although the distribution of errors will be skewed and $E(\hat{y} - y)$ will be greater than zero if $\hat{y} > y$ and vice versa; but this is, indeed, the intended result.

A very few tests have been made to gage the efficacy of this method. A number of random samples of 1000 points each were generated of bivariate normal distributions with unit standard deviation and correlation coefficients of 0.5 and 0.8. The various regression relationships were computed for several values of F (0.1, 0.55, 1.0) as shown in Figs. 2 and 3. The straight lines with the lesser slope are the ordinary least squares regression

lines. The straight lines at 45° to the axes are the "inflated" regression lines. The curved lines are the regression estimates based on the scaling technique.

For the smaller values of the correlation coefficient (Fig. 2, $\rho = 0.5$) the estimates of the deviation of the predictand from its mean made by the scaling technique do not exceed those made by standard regression methods until the independent variable is more than about three standard deviations from its mean. As compared to the inflation methods, the scaling method will predict less extreme values unless the independent variable is four, five, or more standard deviations from its mean (depending on F), a rare event indeed. For $F = 1$, in particular, there are only very slight differences between estimates by normal regression and by the scaling method over a very large range of predictor values.

When the correlation coefficient is larger, however, (Fig. 3, $\rho = 0.8$) the scaling method gives more extreme estimates than standard regression whenever the independent variable is more than about 1.75 standard deviations from its mean, and gives more extreme estimates than even the inflation method for values of the independent variables more than two or three standard deviations of its mean depending on the value of F . It is notable that the estimates by the scaling method are in general close to the standard regression estimates for both values of ρ and for $F = 0.55$ or 1.0 , so long as the independent variable is not very extreme.

Further results are presented in Table 1. When the correlation coefficient is small, the standard error of estimate for the scaling method is only slightly larger than that of the standard method, while the inflation method produces standard errors of estimate which are as large or larger than the predictand standard deviation.

TABLE 1. Comparison of standard errors of estimate and frequency of prediction of extreme events for normal regression, inflation, and scaling methods using various random samples of 1000 points. See text for discussion.

ρ	F	s_y	Standard error of estimate			Frequency of predictions more than 2 standard deviations from mean			
			Normal regression	Inflation	Scaling	Normal regression	Inflation	Scaling	"Observed"
0.5	0.10	0.989	0.873	1.013	0.909	0	40	1	49
0.5	0.55	1.008	0.861	0.970	0.873	0	37	2	47
0.5	1.00	0.989	0.841	0.969	0.852	1	44	1	45
0.8	0.10	1.000	0.584	0.613	2.104	14	47	58	40
0.8	0.55	1.005	0.610	0.716	0.662	17	42	32	47
0.8	1.00	1.009	0.621	0.648	0.638	13	49	20	42

tion. For $\rho=0.8$ and $F=0.1$, the scaling method gives a very large root mean square error estimate. This is almost entirely due to a few extreme cases. For example, one case gave a "prediction" of 15.0 when the "observed" value was 3.9. In this latter case, the "predictions" by the standard and inflation methods were 2.6 and 3.2, respectively. One can ask the question in this case if the prediction of 15.0 wouldn't be "better" than one of 2.6, or even 3.2. Thus, a large standard error of estimate, when due to a few large differences in the sense of overpredicting extremes, should not be interpreted as representing an inadequacy of the method.

On the other hand, this should not be carried to extremes. From the limited evidence available it appears that $F=0.1$ introduces more curvature than is necessary and that a larger value of F is more reasonable. However, it will be necessary to make further tests on less artificial data to allow one to draw any firm conclusions. Also, these tests have been prejudiced against the scaling method in that the relationship between dependent and independent variables was linear. In real cases one can expect non-linear relationships to occur in a form which is as much unlike the curvature introduced by Eq. (1) as it is unlike a straight line, putting both methods at comparable disadvantages.

Also shown in Table 1 is the extent to which "extreme" values, values more than two standard deviations from the mean, are "predicted" by the different techniques. Of necessity the inflation procedure predicts

about as many such extremes as are found in the dependent data. For $\rho=0.5$, normal regression methods produced only one such prediction in 3000 cases, and use of the scaling method improved over this only very slightly. On the other hand, for $\rho=0.8$ there was a substantially larger number of predictions of extremes by scaling than by normal regression, and with $F=0.1$, the scaling method predicted more extremes than inflation.

Only through additional experimentation, analysis, plus the development of suitable criteria and evaluative procedures, can one determine an optimum value of F , or even choose among the various possible procedures (including other functional forms for scaling).

It appears, on the surface at least, that scaling the independent variable in an appropriate way can accomplish some of the goals of the inflation method without all of that method's drawbacks. Further testing and study is required, and seems warranted.

REFERENCES

- Glahn, H. R., and R. A. Allen, 1966: A note concerning the 'inflation' of regression forecasts. *J. Appl. Meteor.*, **5**, 124-126.
- Klein, W. H., C. W. Crockett and C. R. Dunn, 1965: Objective prediction of daily surface temperatures. Tech. Note 10-TDL-1, Techniques Development Laboratory, Weather Bureau, ESSA, 45 pp.
- , B. M. Lewis and I. Enger, 1959: Objective prediction of five-day mean temperature during winter. *J. Meteor.*, **16**, 672-682.