SAMPLE REUSE PROCEDURES FOR CONDITIONAL PREDICTION

by

Seymour Geisser

University of Minnesota

Technical Report No. 346

May, 1979

Sample Reuse Procedures for Conditional Prediction[1]

by
Seymour Geisser
University of Minnesota

I. INTRODUCTION

A data analytic method termed Predictive Sample Reuse (PSR) was introduced by Geisser (1974, 1975) and Stone (1974) which is capable of wide application. It assumes a set of N observations or vectors observed at known covariates and a function, which depends on a set of unknown constants, for predicting potential observables, at known values of the covariates, generated from the underlying process. A set of observables is then deleted and the predictive function formed on the remaining is used to predict the deleted observables. A discrepancy function is defined between the predicted values and the known deleted values which in turn is optimised with respect to the set of unknown constants. The solution for the unknown constants is then utilized in the predictive function for forecasting potential observables.

II. A CONDITIONAL PREDICTION PROBLEM

Suppose N units are observed at the same p time points so that they are represented by p-dimensional vectors X_α, $\alpha = 1,\cdots,N$. Further, an additional vector X_{N+1} is observed at the first $p_1 < p$ points (in actuality, it may be at any of the p_1 out of p points) and the object is to predict (or retrodict past values) the unobserved p_2 values of this partially observed vector. For convenience, we shall assume that we will be dealing with the last p_2 points and define

$$X_\alpha = \begin{matrix} p_1 \\ p_2 \end{matrix} \begin{pmatrix} X_\alpha^{(1)} \\ X_\alpha^{(2)} \end{pmatrix} \qquad (1)$$

$\alpha = 1,\cdots,N+1$ and assume $X_{N+1}^{(2)}$ has not been observed and is to be predicted.

III. PSR APPROACH

Suppose from the first vectors, $X_1,\cdots,X_N$ each at the same p points we generate a predictor of $X_{N+1}^{(2)}$, say $\hat{X}_{(N)}^{(2)}$. Further,

[1]Work supported by NIH grant 1R01 GM25271

suppose another predictor of $X_{N+1}^{(2)}$ is obtained, say $\hat{X}_{N+1}^{(2)}$,

which depends only on the observed $X_{N+1}^{(1)}$. Finally, we combine the two independently calculated predictors into a new predictor

$$\dot{X}_{N+1}^{(2)} = f\left(\hat{X}_{(N)}^{(2)}, \hat{X}_{N+1}^{(2)}; \Omega\right) \tag{2}$$

for $\Omega \varepsilon T, T$ being the specified index set and f an assumed predictive function.

Now let

$$\dot{X}_{\alpha}^{(2)} = f(\hat{X}_{(N-1,\alpha)}, \hat{X}_{\alpha}^{(2)}; \Omega) \tag{3}$$

where $\alpha = 1,\cdots,N$ and $\hat{X}_{(N-1,\alpha)}^{(2)}$ is the predictor for $X_{\alpha}^{(2)}$

based on $X_1,\cdots,X_{\alpha-1},X_{\alpha+1},\cdots,X_N$ and of the same functional form

as $\hat{X}_{(N)}^{(2)}$ and $\hat{X}_{\alpha}^{(2)}$ is the predictor of $X_{\alpha}^{(2)}$ based only on $X_{\alpha}^{(1)}$

and of the same functional form as $\hat{X}_{N+1}^{(2)}$. Further, define an overall discrepancy measure

$$D(\Omega) = D(d_1,d_2,\cdots,d_N) \quad , \tag{4}$$

where $d_{\alpha} = d\left(\dot{X}_{\alpha}^{(2)}, X_{\alpha}^{(2)}\right)$ is some measure of the discrepancy of the predicted value of $X_{\alpha}^{(2)}$, namely $\dot{X}_{\alpha}^{(2)}$ and the actual value, which is then minimized with respect to Ω within its given domain of definition. If $\hat{\Omega}$ is the unique solution and satisfies the constraints, then the final predictor is given as

$$\tilde{X}_{N+1}^{(2)} = f\left(\hat{X}_{(N)}^{(2)}, \hat{X}_{N+1}^{(2)}; \hat{\Omega}\right). \tag{5}$$

An interesting case is

$$\dot{X}_{N+1}^{(2)} = \Omega\, \hat{X}_{(N)}^{(2)} + (I - \Omega)\, \hat{X}_{N+1}^{(2)} \tag{6}$$

where Ω is $p_2 \times p_2$ matrix such that Ω and $I - \Omega$ are both non-negative definite. Define

$$\dot{X}_{\alpha}^{(2)} = \Omega\, \hat{X}_{(N-1,\alpha)}^{(2)} + (I - \Omega)\, \hat{X}_{\alpha}^{(2)} \tag{7}$$

and

$$D(\Omega) = \sum_{\alpha=1}^{N} \left(\dot{X}_\alpha^{(2)} - X_\alpha^{(2)}\right)'\left(\dot{X}_\alpha^{(2)} - X_\alpha^{(2)}\right). \qquad (8)$$

For this situation, a solution may easily be obtained for general forms of $\hat{X}_{(N)}^{(2)}$ and $\hat{X}_{N+1}^{(2)}$, namely

$$\hat{\Omega} = \left[\sum_{\alpha=1}^{N}\left(X_\alpha^{(2)} - \hat{X}_\alpha^{(2)}\right)\left(\hat{X}_{(N-1,\alpha)}^{(2)} - \hat{X}_\alpha^{(2)}\right)'\right]\left[\sum_{\alpha-1}^{q}\left(\hat{X}_{(N-1,\alpha)}^{(2)} - \hat{X}_\alpha^{(2)}\right)\cdot\right.$$

$$\left.\left(\hat{X}_{N-1,\alpha)}^{(2)} - \hat{X}_\alpha^{(2)}\right)'\right]^{-1} \qquad (9)$$

so that

$$\tilde{X}_{N+1}^{(2)} = \hat{\Omega}\,\hat{X}_{(N)}^{(2)} + (I - \hat{\Omega})\,\hat{X}_{N+1}^{(2)}, \qquad (10)$$

provided $\hat{\Omega}$ exists and satisfies the constraints. The simplest way to derive this solution is to use the technique of matix differentiation, c.f. De Waal and Nel (1978).

IV. A SPECIAL CASE

A special case useful in simple growth curve situations where the fitted equations are of the form

$$X = Z_{pxm}\, B_{mx1} = \begin{matrix} p_1 \\ p_2 \end{matrix}\begin{pmatrix} Z_1 \\ Z_2 \end{pmatrix} B \qquad (11)$$

for known Z and $B = (Z'Z)^{-1}Z'\overline{X}$ where $\overline{X} = N^{-1}\sum_{\alpha=1}^{N} X_\alpha$ and $m < p_1$ and p_2 arbitrary. Here, $\hat{X}_{(N)}^{(2)} = Z_2 B$ and $\hat{X}_{N+1}^{(2)} = Z_2(Z_1'Z_1)^{-1}Z_1' X_{N+1}^{(1)}$. A special case of this is given by Geisser (1975) for $m = 2$ and $p_2 = 1$ and Ω a 1×1 scalar. Here

$$Z = \begin{pmatrix} 1 & z_1 \\ \cdot & \\ \cdot & \\ \cdot & \\ 1 & z_p \end{pmatrix} \quad (12)$$

so that one is calculating a convex combination of two predicted values at z_p , one computed from the average of the fitted straight lines of the first N vectors and the other is the predicted value obtained from the line fitted to the $N+1$ vector using the first $p-1$ observed points. Substitution in (9) yields $\hat{\omega}$, the mixing constant.

REFERENCES

De Waal, D. J. and Nel, D. G. (1978), Parametric Multivariate Analysis, University of the Orange Free State, Bloemfontein.

Geisser, S. (1974), "A predictive approach to the random effect model," Biometrika, 61, 1, pp. 101-107.

Geisser, S. (1975), "The predictive sample reuse method with applications," Journal of the American Statistical Association, 70, 350, pp. 320-328.

Stone, M. (1975), "Cross-validatory choice and assessment of statistical predictions," Journal of the Royal Statistical Society, B, 36, 2, pp. 111-147.

SUMMARY

The method of Predictive Sample Reuse is adapted to problems of conditional prediction and applied to a growth curve situation.

SOMMAIRE

La méthode de "Predictive Sample Reuse" s'adapte aux problèmes de la prédiction conditionelle et s'applique à une situation de "Growth Curve."