274

WP: 274

Working Paper

विद्या विनियोगाद्विकासः
IIM
AHMEDABAD

INDIAN INSTITUTE OF MANAGEMENT
AHMEDABAD

PERFECT LEAST SQUARES ESTIMATOR

by

P N Misra
Puneet Handa

W P No. 274
April 1979

INDIAN INSTITUTE OF MANAGEMENT
AHMEDABAD

ABSTRACT

This paper suggests a method of estimation that is the least-squares estimator in the general situation when observations are interdependent or independent. The method is designated as perfect least-squares (PLS) because there is no other method, known so far, that provides lower magnitude of the optimality criterion. The method holds good for data collected according to any sampling method or census method. It is shown in this paper empirically as well as theoretically that PLS estimator scores over OLS and GLS estimators. The method is also extended to simultaneous equation systems. It can be applied straightaway to dynamic models.

PERFECT LEAST SQUARES ESTIMATOR

by

PN Misra & Puneet Handa*

1. Introduction

Most of the methods of estimation are based upon the principle of ordinary least squares (OLS) and its extension, namely, Aitken's generalised least squares (GLS) procedure [1] that accommodates dependence amongst sample observations. GLS procedure is supposed to be the only way to tackle the problem of intra-observational dependence but its quantification has been achieved in certain special cases only. Is GLS really optimal in the sense of minimising error sum of squares in the general case? This question deserves further examination.

The problem can be defined better in context to a general linear model in terms of K variables and n sample observations and expressed as

$$(1.1) \qquad y = X\beta + u$$

where y and u are each nx1 vectors, X is nx(K+1) matrix and β is (K+1)x1 vector. Assuming that the observations are collected according to simple random sampling with replacement so that

$$(1.2) \qquad E(uu') = \sigma^2 I$$

*The authors are Professor and post-graduate student at Indian Institute of Management, Ahmedabad, respectively.

one can obtain best linear unbiased estimator b of β as

(1.3) $b = (X'X)^{-1} X'y$

provided the X matrix is such that

(1.4) $E(X'u) = 0$

or, elements of X are monstochastic and

(1.5) $E(u) = 0$

In all other situations when data are not collected according to simple random sampling with replacement, one may assume that

(1.6) $E(uu') = \Sigma$

and obtain best linear unbiased Aitken's estimator, b_a, as

(1.7) $b_a = (X' \Sigma^{-1} X)^{-1} X' \Sigma^{-1} y$

provided (1.4) or (1.5) hold good. Computation of b_a depends upon the knowledge of Σ which is not estimable completely if only time-series or cross-section data are available.[1] Using estimate of part of Σ matrix is equivalent to misspecification of Σ matrix and in that case properties of Aitken's estimator will be rendered suboptimal.

The properties of the OLS as well as GLS estimator depend upon the assumption that elements of vector u are distributed at each observation point. It is difficult to conceive such a distribution in case of time

[1] An unbiased estimator of all the elements of Σ matrix is provided by Misra [4] in case both time-series as well as cross-section data are available.

series data where observations may relate to entire population or part of a population. This is because population can be defined to be only that time-span over which the parametric vector β remains invariant while in most practical situations β changes over even short time-spans not to speak of long time-spans. It is, therefore, closer to reality in most cases to consider dependence amongst error terms as such without involving their variance-covariance matrix Σ as defined in (1.6) because in that case one is not required to assume that errors are necessarily distributed at each observation point.

These problems can be avoided if we realise that interdependence amongst the elements of u can be considered in functional sense while minimising error sum of squares. Such a consideration will hold good irrespective of whether the data are sample, population or part of population and the sample is with or without replacement because the inherent dependence will be exhibited in the functional dependence used while minimising the error sum of squares. Section 2 of this paper provides an estimator based upon the aforesaid principle. We call this estimator as perfect least-squares (PLS) estimator because it is only this estimator that provides smallest error sum of squares in the general case.

The PLS estimation is extended in Section 3 to several independent or interdependent equations. Its properties are analysed in Section 4. Section 5 contains discussion on the problems involved in quantification of PLS estimator and actually empirical results are reported in Section 6. The last section includes discussion on related issues thrown open by the preceding results. These include analysis of predictive and explanatory powers of linear models besides other issues.

2. The PLS Estimator

Interdependence amongst elements of u can be expressed functionally as

$$(2.1) \qquad u_i = f_i(u_1, \ldots, u_n) \; ; \; i = 1, \ldots, n$$

where functional forms could differ over i. Perfect least-squares estimato of β in (1.1) is obtained by minimising

$$(2.2) \qquad u'u = (y - X\beta)(y - X\beta)'$$

subject to constraint (2.1). In this case, first order condition of minimisation of u'u can be written as

$$(2.3) \qquad \frac{\partial u'u}{\partial \beta} = \sum_i \sum_j \frac{\partial u_i^2}{\partial u_i} \frac{\partial u_i}{\partial u_j} \frac{\partial u_j}{\partial \beta}$$

$$= 2 \frac{\partial u'}{\partial \beta} \frac{\partial u'}{\partial u} u$$

$$= 0$$

where the matrices of partial derivatives are given as

$$(2.4) \qquad \frac{\partial u'}{\partial \beta} = \begin{bmatrix} \frac{\partial u_1}{\partial \beta_1} & \cdots & \frac{\partial u_n}{\partial \beta_1} \\ \vdots & & \vdots \\ \frac{\partial u_1}{\partial \beta_k} & & \frac{\partial u_n}{\partial \beta_k} \end{bmatrix} = -X'$$

$$\frac{\partial u'}{\partial u} = \begin{bmatrix} \frac{\partial u_1}{\partial u_1} & \cdots & \frac{\partial u_n}{\partial u_1} \\ \vdots & & \\ \frac{\partial u_1}{\partial u_n} & \cdots & \frac{\partial u_n}{\partial u_n} \end{bmatrix} = V$$

Using (2.1) and (2.4) we can rewrite (2.3) as

(2.4) $X'VX\beta = X'Vy$

which in turn can be solved for PLS estimator $\hat{\beta}$ of β as

(2.5) $\hat{\beta} = (X'VX)^{-1}X'Vy$

provided X'VX is nonsingular.

Second order condition of minimum of u'u can be seen to hold good provided the matrix

(2.6) $\frac{\partial u'u}{\partial\beta\partial\beta'} = 2X'VX$

is positive definite which depends upon positive definiteness of matrix V. It will be seen later that V can be replaced by several alternative estimates but only positive semidéfinite ones are acceptable in view of aforesaid second order condition.

Considering a general expression to be minimised as

(2.7) $u'Au = (y-X\beta)'A(y-X\beta)$

we observe that the resulting estimator is OLS, GLS or PLS dependirg upon A eing replaced by identity matrix, Σ^{-1}, or V. Let $\tilde{\beta}$ represent estimator of β whenu'Au is minimised, then, u can be estimated as

(2.8) $\hat{u} = y-X\tilde{\beta}$

and the estimated minimum error sum of squares is given by $\hat{u}'A\hat{u}$. This estimate can be used to ascertain empirically as to which one of OLS, GLS and PLS estimators are providing least value of the criterion and the same rule can be used to pick up that estimator of V that provides PLS estimator providing least value of $\hat{u}'V\hat{u}$.

This method can be used straight away to estimate dynamic models of all kinds because it is capable of incorporating in itself the dependence amongst the observations. It can also be used to estimate various time-series models where time is supposed to be the sole causal variable.

3. PLS Estimation of Complete System Model

The procedure of estimation in section 3 can be easily extended to complete system models specified in terms of n observations on each one of endogenous (y) and predetermined (x) variables as

$$(3.1)\quad y_i = Y_i\gamma_i + X_i\beta_i + u_i$$

$$= Z_i\delta_i + u_i$$

$$Z_i = (Y_i \; X_i)$$

$$\delta_i = \begin{bmatrix}\gamma_i\\ \beta_i\end{bmatrix}$$

$$i = 1, \ldots, M$$

where y_i and u_i are (nx1) vector, Y_i is (nxm_i) matrix, γ_i is (m_ix1)vector, X_i is (nxl_i) matrix, β_i is (l_ix1) vector and $n_i = l_i + m_i$. Rewriting the M equations in (3.1) together as

$$(3.2)\quad y^* = Z^*\delta^* + u^*$$

$$y^* = \begin{bmatrix}y_1\\ \vdots\\ y_M\end{bmatrix},\; X^* = \begin{bmatrix}X_1 & \cdots 0\\ \vdots & \\ 0 & X_M\end{bmatrix},\; \delta^* = \begin{bmatrix}\delta_1\\ \vdots\\ \delta_M\end{bmatrix},\; u^* = \begin{bmatrix}u_1\\ \vdots\\ u_M\end{bmatrix}$$

and defining a matrix V*, similar to V in (2.4), as

$$(3.3)\quad V^* = \frac{\delta u^{*\prime}}{\delta u^*}$$

we can obtain PLS estimator $\hat{\beta}^*$ of β^* as

$$(3.4) \quad \hat{\beta}^* = (X^{*'} V^* X^*)^{-1} X^{*'} V^* y^*$$

The matrix V^* incorporates all kinds of dependence amongst the elements of u^* including dependence amongst various equations as well as dependence amongst different observations. It can also be easily seen that the estimatc $\hat{\beta}^*$ in (3.4) can be computed even if number of observations are different in each equation of the system (3.1). The criterion $\hat{u}^{*'} V^* \hat{u}^*$ can be used to make a choice for the appropriate PLS estimator owing to same logic as provided towards the end of Section 3. The estimated residual, $\hat{u}^*$, can be obtained as

$$(3.5) \quad \hat{u}^* = y^* - X^* \hat{\beta}^*$$

4. Properties

In view of discussion in Section 1, we consider the observations to be sample from a given population or the population itself. In case it is interpretted as sample the sample could be simple random sample with or without replacement or any part of the population selected in any other way. The properties of the PLS estimator are proposed to be analysed in case of above mentioned alternatives. We propose to analyse the properties of PLS estimator as given in Section 2 and note that similar results hold good in case of PLS estimator as given in Section 3.

Combining (1.1) with (2.5) we get

$$(4.1) \quad \hat{\beta} = \beta + \beta_e$$
$$\beta_e = (X'VX)^{-1} X'Vu$$

Supposing that u is estimated similar to (2.8) corresponding to PLS estimator $\hat{\beta}$, we find that

$$(4.2) \qquad \beta_e = (X'VX)^{-1}X'V(y-X\hat{\beta}) = 0$$

which shows that so long as $\hat{u}$ is close to u, the estimator $\hat{\beta}$ is close to β. This is possible when the sample covers the entire population. This is also possible when the sample observations are representative enough to provide $\hat{u}$ that is close enough to u. Representative sample in this sense includes those observations that yield the same $\hat{\beta}$ as one would have got by considering the entire population.

Considering the observations to be a sample and assuming that (1.5) holds good, X and V are independent of u, we have

$$(4.3) \qquad E(\beta_e) = (X'VX)^{-1} X'VE(u) = 0$$

Variance-covariance matrix of $\hat{\beta}$ can be obtained as

$$(4.4) \qquad E(\hat{\beta}-\beta)(\hat{\beta}-\beta)' = E(\beta_e\beta_e')$$

so that $\beta_e \beta_e'$ can be used as unbiased estimator of the variance-covariance matrix. Square roots of diagonal elements along $\beta_e \beta_e'$ provide standard errors. In other words, if β_e could be estimated, then, its absolute value provides standard error vector for the estimator $\hat{\beta}$.

Estimated residual sum of squares with PLS procedure can be expressed as

$$(4.5) \qquad \hat{u}'V\hat{u} = y'Vy - y'V\hat{y}$$

so that the total sum of squares in generalised sense can be written as

$$(4.6) \qquad y'Vy = y'V\hat{y} + \hat{u}'V\hat{u}$$

where $y'V\hat{y}$ represents generalised covariation between y and $\hat{y}$. We may define a measure R_p^2 as follows:

$$(4.7) \quad R_p^2 = (y'Vy)^{-1} y'V\hat{y}$$

which is PLS estimator of α coefficient in the model.

$$(4.8) \quad \hat{y}_i = \alpha y_i + v_i \quad ; i = 1, \ldots, n$$

Obviously $\alpha = 1$ for $\hat{y}_i = y_i$ and $\alpha = 0$ if $\hat{y}_i$ and y_i are uncorrelated. The same interpretation follows from (4.6) where $y'V\hat{y}$ is equal to zero on one extreme and $y'vy$ on the other, so that R_p^2 lies between 0 and 1.

Alternatively, one may specify the following regression

$$(4.9) \quad \hat{y}_i = \alpha_0 + \alpha\, y_i + v_i \; ; \; i = 1, \ldots, n$$

and define

$$(4.10) \quad R_f^2 = \text{squared correlation between } \hat{y}_i \text{ and } y_i$$

The measure R_f^2 is same as squared multiple correlation when OLS estimator is used to generate y which excludes the contribution of constant term in explaining y through model (1.1). The measure R_p^2 includes the contribution of constant term also. Conceptually, predictive ability of the model will be better if it scores on R_p^2 criterion because the constant term plays important role in generating forecasts. An attempt to maximise R_f^2 by opting for some specifications may provide distorted estimates because the role of the constant term could get reduced owing to undesirable considerations.

5. Estimation of V.

Computation of PLS estimator depends upon knowledge of matrix V which can be estimated in several ways. An estimate of V can be provided by estimated Σ matrix but in that case the resulting estimator may not possess optimal properties because elements of V and Σ are defined on the basis of different concepts. Alternatively, OLS estimated $\hat{u}$ or any other estimate of u could be used to estimate the elements v_{ij} of V from the following relations.

$$(5.1) \qquad u_i = v_{ij} u_j + e_{ij}$$

It can be readily verified that PLS estimator $\hat{v}_{ij}$ of v_{ij} is given by

$$(5.2) \qquad \hat{v}_{ij} = \frac{u_i}{u_j} = \frac{\partial u_i}{\partial u_j}$$

in view of only one observation available to estimate v_{ij} from model (5.1). The matrix $\hat{V}$ obtained as

$$(5.3) \qquad \hat{V} = ((\hat{v}_{ij}))$$

is PLS estimator of V but one can easily verify that it is singular so that X'VX will also be rendered singular. Therefore this estimator of V is not useful though it provides estimates of all the elements of V.

In actual practice we deal with time-series or cross-section data that are temporally or spatially arranged into some order. As mentioned earlier, blocks of time-series are treated as sample, population or part of population. This provides data in terms of some temporal order which can be utilised in obtaining estimate of matrix V. Similarly, cross-section data possesses some spatial arrangement no matter whether they are sample, population or part of population. This additional information can be used by incorporating the concept of order of dependence in case of time-series as well as cross-section data. If we assume that all the observation units in the first order neighbourhood are homogenous in terms of their interdependence and if we extend the same concept of dependences of higher order we can define the following n regressions:

$$(5.4)\qquad \begin{aligned} u_{i+r} &= v_{or} + v_r\, u_i + e_{i+r} \\ r &= 0, 1, \ldots, n-1 \\ i &= 1, \ldots, n \\ v_{oo} &= 0 \\ v_o &= 1 \\ e_i &= 0 \end{aligned}$$

These regressions express one way causation amongst observation units. One may also define regressions similar to those in (5.4) but expressing the other way causation as

$$(5.5)\qquad \begin{aligned} u_{i-r} &= v^*_{or} + v^*_r\, u_i + e^*_{i-r} \\ r &= 0, 1, \ldots, m-1 \\ i &= 1, \ldots, n \\ v^*_{oo} &\neq 0 \\ v^*_o &= 1 \\ e^*_i &= 0 \end{aligned}$$

Interdependence of types (5.4) and (5.5) can be easily defined in case of cross-section data. It may not be easily explained in case of time-series data in all the situations. In either case what one is interested is observed dependence in either way of causation and that can be estimated by estimating the regressions in (5.4) and (5.5) by using data on error terms. Since data on error terms is difficult to obtain unless one estimates the model according to PLS procedure and PLS estimate cannot be obtained unless V is known, we may follow iterative procedure to overcome this difficulty. We could start with OLS estimated error terms or observations on the dependent variable, y_i, and compute the unknown coefficients in (5.4) and (5.5). Using these estimates we could estimate V because it can be expressed as

$$(5.6) \qquad V = \begin{bmatrix} 1 & v_n & \cdots\cdots\cdots & v_{n-1} \\ \vdots & & & \vdots \\ v^*_{n-1} & v^*_{n-2} & \cdots\cdots & 1 \end{bmatrix}$$

in view of relations (2.4), (5.4) and (5.5).

The iterative process can be carried on by starting with an estimate of V, as above, obtaining PLS estimated errors, using these to estimate V and then repeating the process. The choice of best result can be made by picking out the results corresponding to smallest value of estimated u'Vu. Convergence of estimated V or slope coefficients cannot serve as a criterion of selection because if we start with OLS estimated u then, PLS estimated sampling error defined in (4.1) can be expressed as

$$(5.7) \qquad \beta_e = \hat{\beta} - b$$

which affects estimates of V and β both in the next round and the process continues. Further, these estimates are found to be sensitive to slight changes in data. Therefore consecutive rounds of these estimates are most unlikely to converge. These estimates are PLS if ϱ's are independent.

6. Empirical Results

Empirical results are obtained for an import function, the model and data for which are same as used by Johnston [3, p. 147]. The model is specified as

$$(6.1)\quad y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + u$$

where y represents imports of goods and services, x_1 represents gross U.K. product and x_2 represents ratio of import price to general U.K. output price. The model is estimated according to five methods, namely, OLS, GLS, PLS with V replaced by Σ, PLS when starting with OLS estimated u and PLS when starting with observations on y. We shall designate them as methods 1,2,3,4 and 5, respectively. Of these, method 2 as used in this section requires some explanation. It is assumed that autocorrelations except for first and second order are zero so that an autoregressive model can be specified as

$$(6.2)\quad u_i = \varrho_1 u_{i-1} + \varrho_2 u_{i-2} + \varrho_i$$

Then using Durbin's method [2] we can obtain estimates of ϱ_1 and ϱ_2. These estimates are used in turn to obtain an estimate of Σ matrix where higher order autocorrelations are taken to be zero. This estimate of Σ matrix is used to obtain GLS estimate as defined in (1.7).

The statistics defined in (4.7) and (4.10) are computed in case of all these methods. The main criterion u'Vu is supplemented with corresponding value of u'u to emphasise the magnitude of difference. The standard errors are computed in case of OLS and GLS estimators by using the relevant formulae while those of PLS estimator are computed by using the formula in (4.4). These results are reported in Table 1 below.

Table 1

Alternative Estimates of Model (6.1)

Method	β_0	β_1	β_2	R_p^2	R_f^2	u'u	u'Vu
1	−0.493 (0.058)	1.364 (0.020)	0.114 (0.021)	0.999376	0.93849	7.75	7.75
2	−0.434 (92.19)	1.379 (39.35)	0.064 (23.67)	0.999196	0.93721	12.64	5.75
3	−0.341 (34.086)	1.30 (130.043)	0.037 (3.657)	0.999737	0.93562	8.31	7.96
4	0.129 (0.570)	1.[illegible]5 (0.157)	−0.376 (0.779)	1.00	0.82005	56.01	2.32
5*	0.167 (.990)	1.180 (.461)	−0.293 (.455)	1.00	.849579	23.07	0.31

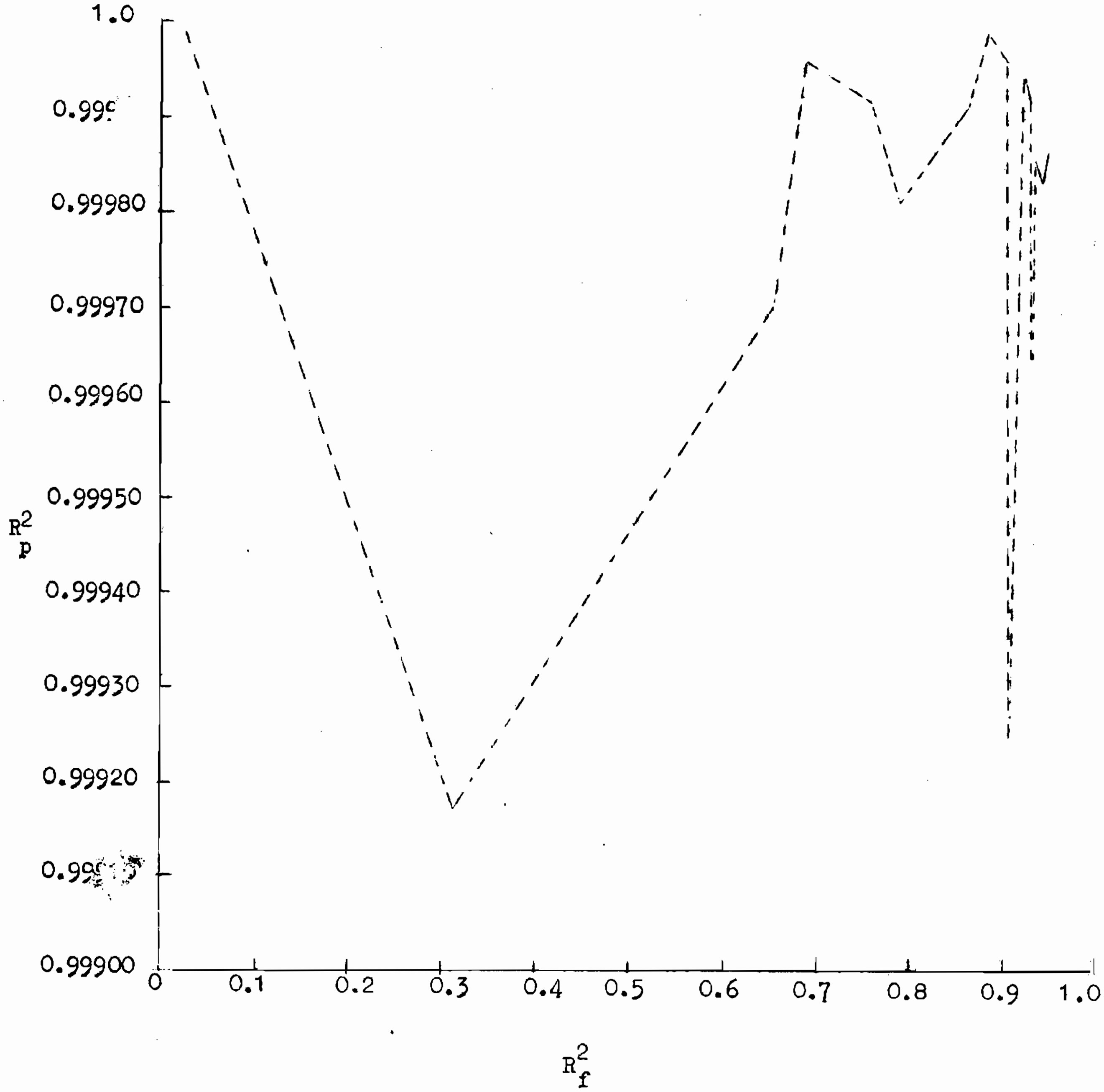
The estimates for u'u and u'Vu are in terms of dimension 10^{-3} where data as mentioned above are adjusted by multiplying each y, x_1, x_2 by 10^{-2}. The figures in brackets are standard errors. Results relating to methods 4 and 5 correspond to rounds 13 and 7, respectively from a total run of 40 and 38 rounds respectively.

Considering the basic criterion, namely, u'Vu, the PLS method as obtained by using observations on y to obtain starting estimate of V turns out to be best. The method provides negative sign of coefficient of x_2 which is desirable in view of economic theory whereas both OLS and GLS estimates of this coefficient are positive. An attempt to use Σ in place of V to obtain PLS estimator leads result that is worse than OLS estimator. This leads to suggest that method 5 is better than others from economic as well as optimality considerations. Another interesting point that comes out from the above exercise is that empirically GLS estimators could be inefficient as compared to OLS estimators though theoretically the reverse holds good.

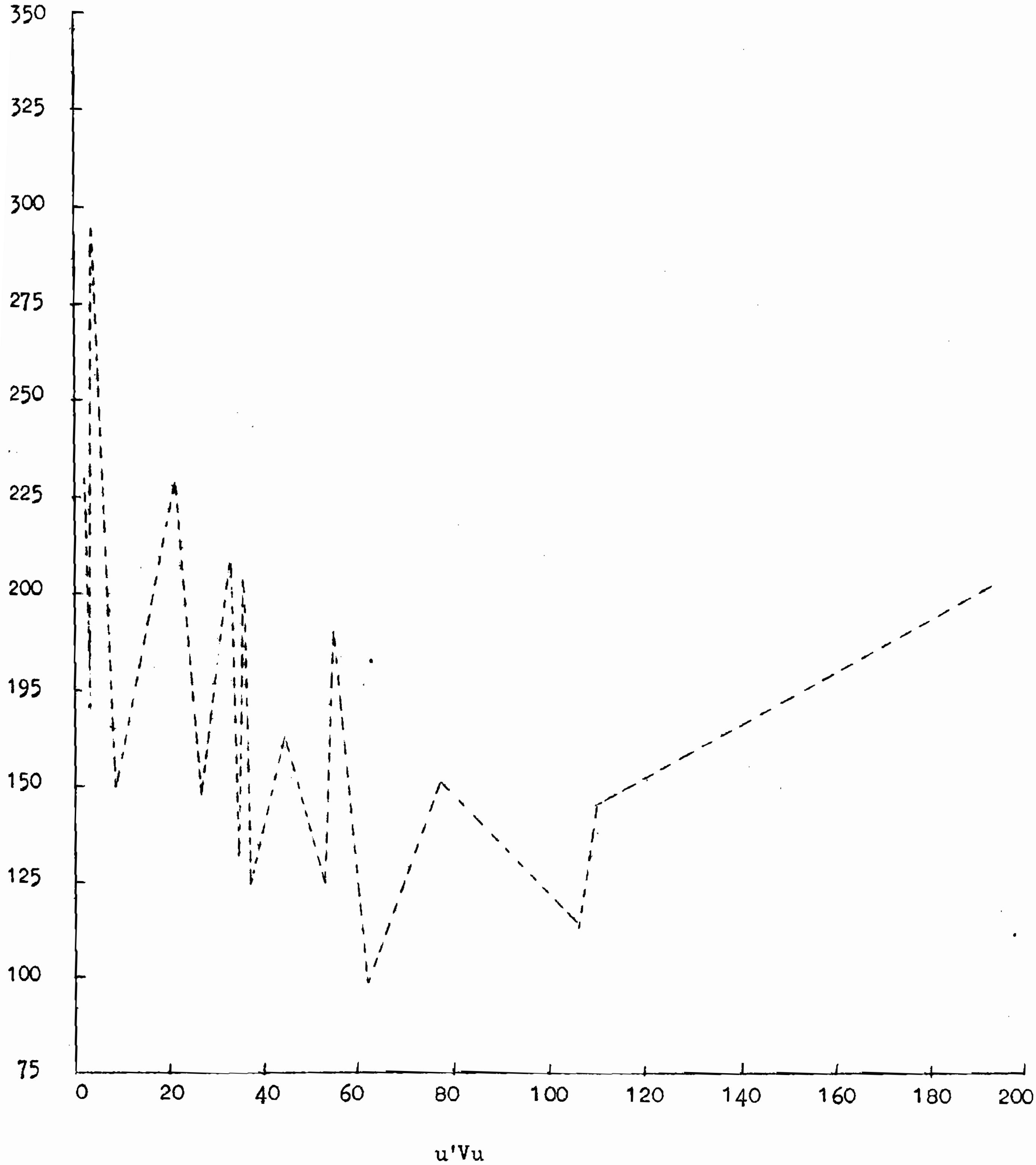
Estimates of the statistic R_p^2 and R_f^2 are available for various rounds in case of method 4 and 5. We use these results corresponding to method 5 with positive valued u'Vu and plot them on graph 1.

The scatter of the points on this graph suggests that neither R_p^2 nor R_f^2 can be used as the sole criterion of deciding predictive power of a model. However, R_p^2 coupled with u'Vu can be used to judge the predictive efficacy of a model and this is supported by the results in Table 1.

Empirical results of u'Vu and u'u are available for various rounds in case of methods 4 and 5. We use the results corresponding to method 5 for positive values of u'Vu and plot them as below in graph 2.


1.0
0.999
0.99980
0.99970
0.99960
0.99950
0.99940
0.99930
0.99920
0.99900
R^2_p
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1.0
R^2_f

Graph 1


350
325
300
275
250
225
200
195
150
125
100
75
0
20
40
60
80
100
120
140
160
180
200
u'Vu

Graph 2

The scatter of points in the above graph suggests lack of any correspondence between the two criteria. Therefore any attempt to use the criterion u'u without having any evidence that V = I is most likely to lead to quite sub-optimal results and this should be avoided in all the empirical studies.

REFERENCES

1. Aitken, A.C., "On Least-squares and Linear Combination of Observations," Proc. Royal Soc., Edinburgh, Vol.55, pp 42-48, 1934.

2. Durbin, J., "Estimation of Parameters in Time-series Regression Models," Journal of Royal Statist. Soc., Series B, Vol.22, pp 139-153, 1960

3. Johnston, J., Econometric Methods, Second Edition, McGraw Hill Kogakusha Ltd., Tokyo, Japan, 1972.

4. Misra, P.N., "Pooling of Time-series and Cross-section Data," Sankhya, Series B, Vol. 34, Part 4, pp 385-394, 1972.