COMMUN. STATIST.-SIMULA. COMPUTA., B10(3), 303-313 (1981)

SMALL-SAMPLE BEHAVIOR OF WEIGHTED LEAST SQUARES IN EXPERIMENTAL DESIGN APPLICATIONS

Jack Kleijnen
Renée Brent
Rien Brouwers

Department of Business and Economics
Tilburg University (Katholieke Hogeschool Tilburg)
Tilburg, Netherlands.

Key Words and Phrases: generalized least squares; experimental error; Monte Carlo.

ABSTRACT

In experimental design applications unbiased estimators s_i^2 of the variances σ_i^2 are possible. These estimators may be used in Weighted Least Squares (WLS) when estimating the parameters β. The resulting small-sample behavior is investigated in a Monte Carlo experiment. This experiment shows that an asymptotically valid covariance formula can be used if s_i^2 is based on, say, at least 5 observations. The WLS estimator based on estimators s_i^2 gives more accurate estimators of β, provided the σ_i^2 differ by a factor, say, 10.

1. INTRODUCTION

Generalized Least Squares (GLS) is a popular technique in econometrics, a discipline that can be characterized as follows: (i) In econometrics (and in the social sciences in general) the explanatory or independent variables x can not be fixed by the researcher. These variables are fixed by the environment; the investigator can only observe the variables.
(ii) Economic and statistical reasoning leads to rejection of the classical assumption that stipulates independent errors $\underset{\sim}{e}$ with constant but unknown variance σ^2:

$$\underset{\sim}{\Omega}_e = \sigma^2 \underset{\sim}{I} \tag{1.1}$$

where $\underset{\sim}{\Omega}$ denotes a covariance matrix and $\underset{\sim}{I}$ the identity matrix. A consequence of characteristic (i) is that replication of specific experimental conditions is impossible. Hence $\underset{\sim}{\Omega}_e$ is estimated from the residuals $(\underset{\sim}{y}-\hat{\underset{\sim}{y}})$.

In experimental design applications the variables x are fixed by the scientist, either in a real-life experiment or in a simulation experiment. Each experimental condition i can be replicated resulting in the unbiased estimators

$$s_i^2 = \sum_{j=1}^{m_i} (y_{ij}-\bar{y}_i)^2/(m_i-1) \quad (i = 1,\dots,n) \tag{1.2}$$

In general it is realistic to replace eq. (1.1) by

$$\underset{\sim}{\Omega}_e = \underset{\sim}{D} \equiv \begin{bmatrix} \sigma_1^2 & & & \\ & \sigma_2^2 & & \\ & & \ddots & \\ & & & \sigma_n^2 \end{bmatrix} \tag{1.3}$$

where $\underset{\sim}{D}$ denotes a diagonal matrix, i.e. the n experiments are assumed to be independent. In a simulation experiment - as opposed to a real-life experiment - independence is guaranteed by the use of different random number streams.

In experimental design practice Ordinary Least Squares (OLS) is standard. Let us briefly review OLS versus GLS. Consider the model linear in the parameters β:

$$y = X\beta + e \quad (1.4)$$

The technique for estimating β depends on the experimenter's criterion or loss function. OLS minimizes $\Sigma(y_i - \hat{y}_i)^2$, resulting in

$$\hat{\beta} = (X'X)^{-1} X'y \quad (1.5)$$

GLS - more specifically Weighted Least Squares (WLS) - minimizes $\Sigma\omega_i(y_i - \hat{y}_i)^2$ with weights $\omega_i = 1/\sigma_i^2$ yielding

$$\beta^* = (X'D^{-1}X)^{-1} X'D^{-1}y \quad (1.6)$$

If the classical assumptions of eq. (1.1) hold, then OLS yields best linear unbiased estimators (BLUE) $\hat{\beta}$ with

$$\Omega_{\hat{\beta}} = \sigma^2(X'X)^{-1} \quad (1.7)$$

If, however, the more realistic and general assumption (1.3) holds, then GLS results in BLUE β with

$$\Omega_{\beta^*} = (X'D^{-1}X)^{-1} \quad (1.8)$$

The practical problem addressed in this note is as follows: The GLS estimators β^* assume non-constant but known variances σ_i^2. In practice the nuisance parameters σ_i^2 are unknown. Estimating σ_i^2 results in the estimator $\hat{D}$ which upon substitution into eq. (1.6) yields Estimated GLS (EGLS) estimators $\hat{\beta}^*$. As Schmidt (1976, p. 71) shows this new estimator has the same asymptotic distribution as the regular GLS estimator and (under certain mild technical conditions) remains unbiased. Unfortunately its small-sample behavior is unknown. It is our purpose to examine the small-sample behavior in experimental design situations, as opposed to econometric situations. For that study we use the Monte Carlo method.

We emphazise that even if the classical assumption (1.1) does not hold, OLS can still be used as a mathematical technique for estimating the parameters β. However, the statistical properties of the OLS estimator become different: $\hat{\beta}$ is no longer known to be BLUE and the covariance matrix (1.7) is replaced by

$$\Omega_{\hat{\beta}} = W \, \Omega_e \, W' \tag{1.9}$$

using the shorthand notation

$$W \equiv (X'X)^{-1}X' \tag{1.10}$$

The OLS estimator does remain unbiased, because eqs. (1.5), (1.4) and (1.10) yield

$$\mathcal{E}(\hat{\beta}) = W \, \mathcal{E}(y) = \beta \tag{1.11}$$

Note that if X is ill-conditioned (nearly singular) then special numerical techniques are needed; see Wampler (1979).

2. MONTE CARLO EXPERIMENT.

In section 1 we postulated that in experimental design applications each experimental condition i ($i = 1,\ldots,n$) yields a variance estimator s_i^2 defined in eq. (1.2). In our Monte Carlo experiment we further assume (for convenience) that each condition i is replicated the same number of times: $m_i = m$. In a simulation experiment with, say, a queuing system m denotes the number of independent subruns into which the simulation run is divided; see

Kleijnen (1979).[1] In Monte Carlo experiments m denotes the number of replications using different random number streams. Since another classical assumption namely normality, is not examined in this study, e_{ij} is sampled[2] m times from $N(0,\sigma_i^2)$ and the average

$$\bar{e}_i = \sum_{j=1}^{m} e_{ij}/m \tag{2.1}$$

is computed so that

$$\bar{y}_i = \sum_{k=1}^{q} \beta_k x_{ik} + \bar{e}_i \tag{2.2}$$

Hence in the notation of section 1 we have $\underset{\sim}{y}' = (\bar{y}_1,\ldots,\bar{y}_n)$.

The output of the Monte Carlo experiment consists of estimates of the covariance matrices of $\hat{\underset{\sim}{\beta}}$ and $\hat{\underset{\sim}{\beta}}^*$. Moreover the average values of $\hat{\underset{\sim}{\beta}}$ and $\hat{\underset{\sim}{\beta}}^*$ are computed. The average $\hat{\beta}$ can be used to verify the correctness of our program: OLS should yield unbiased estimates; see eq. (1.11). The other outputs will be presented in the next section.

The input parameters of our Monte Carlo experiment are: m, n, q, $\underset{\sim}{\beta}$, $\underset{\sim}{X}$, $\underset{\sim}{D}$, and the number of times the Monte Carlo experiment is repeated, say, M. In a trial experiment we used M = 250 and could reach significant conclusions. To save computer time we reduced M to 150 and as we shall see significant conclusions can still be reached. The emphasis is on how differences in the variances σ_i^2 which occur in $\underset{\sim}{D}$, affect OLS and EGLS. Hence one extreme is constant variances: $\sigma_i^2 = \sigma^2$. The other extreme is provided by a case-study (a harbor simulation) where the maximum and minimum s_i^2 differed by a factor 1,456; see Kleijnen et al. (1979). Some intermediate cases will be shown in the tables. The same case-study provided a specific orthogonal X-matrix and a specific $\underset{\sim}{\beta}$ with n = 16 and q = 13.[3] Asymptotic results for EGLS are examined by varying the degrees of freedom (m-1) on which s_i^2 is based. An absolute minimum is m = 2. In the case-study we had m = 9. Several more values are shown in the tables. All Monte Carlo results are for orthogonal $\underset{\sim}{X}$.

3. MONTE CARLO RESULTS

The Monte Carlo experiment is analysed as follows:[4)]
(1) We test whether EGLS yields unbiased estimators, as is known to occur when $e_{ij} \sim N(0,\sigma_i^2)$. Therefore we compute per parameter β:

$$t_{M-1} = \frac{\bar{\hat{\beta}}^{*} - \beta}{\{\hat{\mathrm{var}}(\hat{\beta}^{*})/M\}^{1/2}} \tag{3.1}$$

Since M = 150 t_{M-1} can be replaced by $z \sim N(0,1)$. Among the 141 observations[5)] on t only one t-value is significant: t = 2.88 where as $z^{\alpha/2} = 2.60$ for $\alpha = 0.01$. If the null-hypothesis holds then the expected number of significant values is (0.01) (141) = 1.41. So EGLS indeed yields unbiased estimators $\hat{\underset{\sim}{\beta}}^{*}$.
(2) We test whether the asymptotic covariance matrix (1.8) is a valid approximation. Since in practice an experimenter uses only the main diagonal elements ω_{ii} of eq. (1.8) we compute for each of the q parameters β:

$$\chi^2_{M-1} = \frac{\hat{\mathrm{var}}(\hat{\beta}^{*})}{\omega_{ii}} \tag{3.2}$$

In Table I the factor "heterogeneity" denotes

$$H = \frac{\sigma^2_{max} - \sigma^2_{min}}{\sigma^2_{min}} \tag{3.3}$$

where σ^2_{max} and σ^2_{min} are the *maximum* and *minimum* elements on the diagonal of $\underset{\sim}{D}$. This factor will become relevant in step (3). Table I displays only the maximum of the q χ^2-values. (A "per comparison" error rate of $\alpha = 0.01$ corresponds to a familywise error rate of $q\alpha$; see Kleijnen, 1975, pp. 526-531.) Note that some (non-displayed) χ^2-values are smaller than 1, i.e., eq. (1.8) does not

TABLE I

Adequacy of asymptotic variance formula (1.8)

Case 1: n = 16, q = 13

Heterogeneity H

	0	11.84			1,455.69		
m	9	2	9	25	2	9	25
max χ^2	1.26	1.90*	1.22	1.25	1.88*	1.25	1.22

Case 2: n = 8, q = 4

Heterogeneity H

	0		10.83			1,455.69		
m	9	25	2	9	25	2	9	25
max χ^2	1.24	1.17	3.30*	1.24	1.17	7.79*	1.21	1.20

Case 3: n = 4, q = 3

Heterogeneity H

	10.38				1,289.15	
m	4	5	9	25	4	5
max χ^2	1.43*	1.20	1.17	1.19	1.55*	1.18

systematically underestimate var($\hat{\beta}^*$). Only if σ_i^2 is estimated from _extremely few_ observations like m = 2, the asymptotic formula _underestimates_ the true variance of $\hat{\beta}^*$ significantly. For a small X-matrix the effect of m is further examined: case 3. The underestimation becomes insignificant even for m as small as 5; since in practical simulation experiments m is always higher than 5, we recommend the use of the asymptotic formula (1.8) in simulation studies.

TABLE II

Efficiency of EGLS versus OLS

Case 1: n = 16, q = 13

Heterogeneity H

	0	11.84			1,455.69		
m	9	2	9	25	2	9	25
min χ^2	0.82	0.92	0.62*	0.72*	0.10*	0.07*	0.07*
max χ^2	1.26	1.88*	1.18	1.13	1.51*	1.00	0.97

Case 2: n = 8, q = 4

Heterogeneity H

	0		10.83			1,455.69		
m	9	25	2	9	25	2	9	25
min χ^2	0.92	0.96	1.34	0.58*	0.52*	0.16*	0.05*	0.06*
max χ^2	1.24	1.17	1.89*	0.70	0.67	1.37*	0.21	0.21

Case 3: n = 4, q = 3

Heterogeneity H

	10.38				1,289.15	
m	4	5	9	25	4	5
min χ^2	0.57*	0.55*	0.63*	0.47*	0.17*	0.19*
max χ^2	1.17	0.89	0.85	0.88	1.06	0.81

(3) We test whether EGLS is "better" than OLS. Therefore we compare the Monte Carlo estimates of the variances of $\hat{\beta}^*$ with the known variances σ_{ii} of $\hat{\beta}$ (see eq. (1.9)):[6)]

$$\chi^2_{M-1} = \frac{\hat{\text{var}}(\hat{\beta}^*)}{\sigma_{ii}} \tag{3.4}$$

Table II displays not all q values of χ^2 but only the maximum and minimum value. A significant maximum χ^2 and insignificant minimum χ^2 means that OLS is better; a significant minimum χ^2 and insignificant maximum χ^2 means that EGLS is better. Whether OLS is worse than EGLS depends on the degree of heterogeneity of variance, measured by H in eq. (3.3).

(i) In case of perfect homoscedasticity (H = 0) EGLS estimates $\underset{\sim}{D}$ whereas OLS correctly assumes that $\underset{\sim}{D} = \sigma^2\underset{\sim}{I}$. Hence OLS gives better results but, as Table II shows, not significantly better. Consequently if we knew that the assumption of constant variances holds then we could recommend OLS.

(ii) For strong heteroscedasticity (H > 1,000) all our tests (remember note 6) show that OLS is inferior, provided m > 2 so that reasonable estimates of σ_i^2 are possible.

(iii) In case of "intermediate" heterogeneity (H ≈ 10) EGLS is still better than OLS provided m > 2. As expected the superiority of EGLS is smaller than in case of strong heterogeneity. (For case 1 the tests of note 6 do not give significant results for intermediate heterogeneity.)

Note that if n = q then OLS and EGLS become identical.[7)]

4. PRACTICAL CONCLUSIONS

(1) In real-life experiments the individual variances σ_i^2 are estimated from only a few replications (m) because of the effort and practical difficulties involved. For "small" m (m between 1 and, say, 5) we recommend OLS for the estimation of the parameters $\underset{\sim}{\beta}$. However, the standard errors of $\hat{\underset{\sim}{\beta}}$ should incorporate the variance estimators s_i^2 using eq. (1.9) which replaces the standard formula, eq. (1.7).

(2) In simulation experiments replication is easy so that m is "large", say, $m > 5$. If the estimators s_i^2 indicate heterogeneity, say, $s_{max}^2/s_{min}^2 > 10$ then we recommend WLS; else we recommend OLS together with eq. (1.9).

NOTES

1. In renewal analysis of simulation m must be "large", say, $m > 100$.
2. The random number generator on our ICL 2900 turned out to be undocumented, i.e., it is a multiplicative congruential generator with unknown parameters. Normal variates are generated through a subroutine presumed to be the Box-Muller transformation.
3. $\underset{\sim}{\beta}'$ = (-1.420, -0.769, 13.440, -11.508, 3.500, -1.375, 140.918 15.391, 0.046, 281.098, 21.250, 11.875, -49.483). $\underset{\sim}{X}$ was defined by the experimental design "generators" 1 = 56 and 3 = 45.
4. We also verified whether the program indeed yields unbiased OLS estimators, computing per parameter β

   $$z = (\bar{\hat{\beta}}-\beta)/\{var(\hat{\beta})/M\}^{1/2}$$

   where $var(\hat{\beta})$ follows from eq. (1.9) so that $z \sim N(0,1)$. We further verified whether the estimated variances agree with eq. (1.9), using the χ^2 statistic of eq.(3.2) . At $\alpha = 0.01$ none of the many z or χ^2 values is significant.
5. Each combination of a β parameter and an m value yields one t observation: $141 = 7 \times 13 + 8 \times 4 + 6 \times 3$.
6. We also compared the estimated variances of EGLS with the <u>estimated</u> variances of OLS using an F-statistic. Moreover we computed the eigenvalues of $\hat{\underset{\sim}{\Omega}}_{\hat{\beta}}-\hat{\underset{\sim}{\Omega}}_{\hat{\beta}^*}$, and the difference between the traces of these two covariance matrices. All tests yield the same qualitative conclusions.
7. If $n = q$ then X^{-1} exists. Hence $\hat{\underset{\sim}{\beta}}$ and $\hat{\underset{\sim}{\beta}}^*$ reduce to $\underset{\sim}{X}^{-1}\underset{\sim}{y}$.

BIBLIOGRAPHY

Kleijnen, J.P.C., Statistical Techniques in Simulation (In two volumes). Marcel Dekker, Inc., New York, 1974/1975.

Kleijnen, J.P.C., The role of statistical methodology in simulation. In: Methodology in Systems Modelling and Simulation, edited by B. Zeigler, M.S. Elzas, G.J. Klir and T.I. Ören, North-Holland Publishing Company, Amsterdam, 1979.

Kleijnen, J.P.C., A.J. van den Burg and R.T. van der Ham, Generalization of simulation results: practicality of statistical methods. European Journal Operational Research, 3, 1979, pp. 50-64.

Schmidt, P., Econometrics. Marcel Dekker, Inc., New York, 1976.

Wampler, R.H., Solution to weighted least squares problems by modified Gram-Schmidt with iterative refinement. ACM Transactions on Mathematical Software, 5, no. 4, Dec. 1979, pp. 457-465.

Received February, 1980; Revised July, 1980.

Recommended by Ralph A. Bradley, Florida State University, Tallahassee, FL.

Refereed Anonymously.