FEW

faculteit der economische wetenschappen

RESEARCH MEMORANDUM

TILBURG UNIVERSITY

DEPARTMENT OF ECONOMICS

Postbus 90135 - 5000 LE Tilburg
Netherlands

REGRESSION ANALYSIS FOR
SIMULATION PRACTITIONERS

Jack P.C. Kleijnen

CONTENTS

REGRESSION ANALYSIS FOR SIMULATION PRACTITIONERS

Jack P. C. Kleijnen

Katholieke Hogeschool/Indiana University*

ABSTRACT

Based on elementary regression analysis as found in standard
textbooks and computer software packages this tutorial presents some
simple extensions useful in the analysis of simulation experiments.
In simulation one usually has variance estimates (standard errors)
available. These estimates often conflict with the assumption of
constant variances maintained in elementary regression analysis.
Therefore two options are available: (1) Switch to Generalized Least
Squares, (2) Continue to use Ordinary Least Squares. The consequences
of both approaches are surveyed. How to test the model's adequacy is
discussed in detail.

1. INTRODUCTION

In most practical and theoretical simulation studies the ex-
perimenter obtains an estimate not only of the mean system response
(e.g. queuing time) but also of the standard errors of this estimate.
The standard errors s_i (i=i, ...,N) of the responses for N different
system configurations often show large differences, and hence the
assumption of constant variance $\sigma_i^2 = \sigma^2$ obviously does not hold. For

---

an illustration I refer to the case-study discussed later (see
Table 2) in which the estimated variances s_i^2 vary between 64 and
93,228. Many more examples could be given.[1] Observe that the foot-
notes make it possible to read this tutorial at two levels: as an
elementary survey using a minimum of formulas or as a survey includ-
ing technical details and additional references.

It has become more and more accepted to analyse the outputs
of a simulation experiment, applying techniques like Analysis of
Variance (ANOVA)[2] and regression analysis; see Weeks & Fryer
(1976) for a practical example and see Kleijnen (1979) for additional
references. However, in virtually all practical applications
simulationists have assumed a constant variance σ^2. Such a practice
is stimulated by the availability of standard computer packages,
based on the constant variance assumption[3].

Remember that Ordinary Least Squares (OLS) refers to a strictly
mathematical (i.e. non-statistical) criterion: minimize the sum of
squared deviations. If the standard statistical assumptions of
normally and independently distributed (NID) errors e with constant
variance σ^2 and zero expectation

$$e_i \sim NID\ (0,\sigma^2) \tag{1.1}$$

are introduced, then the OLS estimator is known to be BLUE: best
linear unbiased estimator, "best" meaning minimum variance. If the
variances σ_i^2 are unequal, the simulationist may stick to the OLS
algorithm but he has to be aware of the different values for the
standard errors of the traditional OLS estimators,

$$\hat{\underset{\sim}{\beta}} = (\underset{\sim}{X}'.\underset{\sim}{X})^{-1}\ \underset{\sim}{X}'.\underset{\sim}{y} \tag{1.2}$$

where $\underset{\sim}{X}$ and $\underset{\sim}{y}$ are the standard regression symbols. Using elementary statistics the variances (and covariances) of $\hat{\underset{\sim}{\beta}}$ will be investigated.

The simulation practitioner may also remember that if the standard assumptions in eq. (1.1) do <u>not</u> hold, then a BLUE results when Generalized Least Squares (GLS) is applied:

$$\widetilde{\underset{\sim}{\beta}} = (\underset{\sim}{X}' \cdot \underset{\sim}{\Omega}^{-1} \cdot \underset{\sim}{X})^{-1} \underset{\sim}{X}' \cdot \underset{\sim}{\Omega}^{-1} \cdot \underset{\sim}{y} \qquad (1.3)$$

where $\underset{\sim}{\Omega}$ denotes the "covariance matrix" i.e., the matrix of variances and covariances of e_i (or equivalently y). However, in practice $\underset{\sim}{\Omega}$ is unknown and has to be estimated. Substitution of the estimator $\hat{\underset{\sim}{\Omega}}$ into eq. (1.3) yields an estimator, say $\hat{\underset{\sim}{\beta}}$, with unknown <u>small-sample</u> properties! Below the results of a small scale Monte Carlo experiment are reported, which may be used as a preliminary guideline.

I have applied both techniques, OLS and GLS, to a practical simulation experiment; see Kleijnen et al. (1979). The results will be summarized in section 6.

2. ORDINARY LEAST SQUARES AND HETEROGENOUS VARIANCES

As mentioned above, OLS can still be used even if the standard assumptions of eq. (1.1) are violated. To derive the standard errors of the corresponding estimators $\hat{\underset{\sim}{\beta}}$, a result presented in e.g. Scheffé (1964, p. 8) is needed. Consider a vector of stochastic variables, say $\underset{\sim}{Y}$, with covariance matrix $\underset{\sim}{\Omega}_1$. Next introduce a linear transformation of $\underset{\sim}{Y}_1$:

$$\underset{\sim}{Y}_2 = \underset{\sim}{A} \cdot \underset{\sim}{Y}_1 \qquad (2.1)$$

Then $\underset{\sim}{Y}_2$'s covariance matrix can be proven to be

$$\underset{\sim}{\Omega}_2 = \underset{\sim}{A} \cdot \underset{\sim}{\Omega}_1 \underset{\sim}{A}'$$
(2.2)

Applying this result to eq. (1.2), defining for convenience

$$\underset{\sim}{W} = (\underset{\sim}{X}'.\underset{\sim}{X})^{-1} \underset{\sim}{X}'$$
(2.3)

results in the covariance matrix of $\hat{\underset{\sim}{\beta}}$:

$$\underset{\sim}{\Omega}_{\hat{\beta}} = \underset{\sim}{W}.\underset{\sim}{\Omega}.\underset{\sim}{W}'$$
(2.4)

Eq. (2.4) together with eq. (2.3) looks quite complicated. Nevertheless an estimator $\hat{\underset{\sim}{\Omega}}_{\hat{\beta}}$ can be easily computed through a computer program that reads the values of the independent variables $\underset{\sim}{X}$ and the estimator $\hat{\underset{\sim}{\Omega}}$. Obviously the OLS estimator remains unbiased. [4]

In *simulation* the observations $\underset{\sim}{y}$ can be made strictly independent through the use of different random numbers per simulation run. Hence $\underset{\sim}{\Omega}$ reduces to a diagonal matrix, say $\underset{\sim}{D}$. As mentioned in section 1, its diagonal elements σ_i^2 will not be constant in general. In the simulation of steady-state behavior, runs might be continued so long that each run yields the same *estimated* variance. In practice such an approach is not popular.

If the standard assumptions of eq. (1.1) held, i.e. $\underset{\sim}{\Omega} = \sigma.\underset{\sim}{I}$, then eq. (2.4) would reduce to a familiar formula:

$$\underset{\sim}{\Omega}_{\hat{\beta}} = \sigma^2 .(\underset{\sim}{X}'.\underset{\sim}{X})^{-1}$$
(2.5)

In practice eq. (2.5) is applied by simulationists relying on standard software. Usually the common variance σ^2 in eq. (2.5) is estimated through the Mean Squared Residuals (MSR):

$$MSR = \sum_{1}^{N} (y_i - \hat{y}_i)^2 /(N-q)$$
(2.6)

where q denotes the number of estimated parameters, i.e., $\hat{\underset{\sim}{\beta}}$ is a vector of q elements . The MSR has only (N-q) degrees of freedom

(d.f.) whereas in simulation each run provides an estimator s_i^2 with d_i degrees of freedom when the total run i is divided into $(d_i + 1)$ independent subruns.[5] If a common variance were assumed, the N runs could be combined to yield a pooled estimator of σ^2 with Σd_i degrees of freedom. Hence I would recommend that simulationists use the information on the standard errors to get a more precise estimator of σ^2, if a common variance assumption is maintained.

Simulationists familiar with experimental design should note that even with an orthogenal X-matrix, the OLS estimators remain correlated when the observations have heterogenous variances.[6]

Summarizing so far, if the constant variance assumption is not maintained then eqs. (2.5) and (2.6) are replaced by eq. (2.4) where $\underset{\sim}{\Omega}$ is estimated from the N individual simulation runs, and $\underset{\sim}{\Omega}$ becomes the diagonal matrix $\underset{\sim}{D}$ with elements s_i^2, each s_i^2 having d_i degrees of freedom.[7] Therefore, the significance of an estimated regression parameter $\hat{\beta}_j$ (j = 1, ...,q) can be tested through the Student t-test:

$$t_d = \frac{\hat{\beta}_j - \beta_j^0}{\sqrt{\hat{\text{var}} (\hat{\beta}_j)}} \qquad (2.7)$$

Here β_j^0 denotes the hypothesized value, usually zero. The denominator follows from the main diagonal of $\hat{\Omega}_\beta$. The index d denoted the d.f. of t. In simulation s_i^2 has so many d.f. that the t-distribution can be replaced by the standard normal distribution.[8]

If the postulated value β_j^0 is accepted then the regression model's remaining parameters $\beta_{j'}$, $(j' \neq j)$ can be reestimated. With an orthogonal X-matrix, the reestimated parameters, will remain unchanged.

3. GENERALIZED LEAST SQUARES

The GLS estimator $\tilde{\beta}$ was defined in eq. (1.3). For independent observations $\underset{\sim}{\Omega}$ reduces to the diagonal matrix $\underset{\sim}{D}$ and GLS can be simplified to "weighted least squares", the weight for observation y_i being inversely proportional to its variance σ_i^2; see e.g. Pritchard & Bacon (1977). The covariance matrix of the GLS estimator is known to be

$$\underset{\sim}{\Omega}_{\tilde{\beta}} = (\underset{\sim}{X}' \cdot \underset{\sim}{\Omega}^{-1} \cdot \underset{\sim}{X})^{-1} \tag{3.1}$$

Since in practice $\underset{\sim}{\Omega}$ is unknown, two options are available:

(1) Estimate $\underset{\sim}{\Omega}$ and substitute the estimator $\hat{\underset{\sim}{\Omega}}$ into the GLS equation (1.3). As Schmidt (1976, pp. 71-72) shows the resulting estimator has the same asumptotic distribution as the regular GLS estimator and remains unbiased (under mild technical conditions). Unfortunately, its small sample behavior remains unknown!

(2) Use OLS even when the classical assumptions of eq. (1.1) are violated, using eqs. (2.4) and (2.7).[9]

In appendix 1 a small Monte Carlo experiment is presented. Based on this preliminary experiment I conjecture:

(a) GLS with estimated covariance matrix $\underset{\sim}{\Omega}$ gives point estimators with smaller variances than OLS estimators. This result seems intuitively acceptable because OLS yields BLUE only if the variances σ_i^2 were constant; the "estimated GLS" tries to incorporate the information s_i^2 on the actual variances σ_i^2.

(b) For the "estimated GLS" estimators $\hat{\beta}$ the standard errors might still be computed through eq. (3.1), a formula - strictly speaking - valid for known Ω or for "large" samples. Intuitively, replacing Ω by its estimator $\hat{\Omega}$ increases the variance compared to eq. (3.1).

Because of (a) and (b) I would suggest to use GLS. A case-study employing both OLS and GLS, will be presented later.

4. VALIDATION OF REGRESSION METAMODEL

The metamodel should explain how the more complicated simulation model's output y reacts to changes in the simulation model's input factors x_1 through x_k ($k \geq 1$). The experimental design fixes x_{i1} through x_{ik} with i = 1, ...,N. The metamodel may further include interaction terms like $x_{i1} x_{ik}$, quadratic terms like x_{i1}^2, etc. which are completely determined by the choice of the design; see Kleijnen (1979). Deciding which interactions to include in X specifies the form of the metamodel, linear in its parameters β:

$$y = X.\beta + e \qquad (4.1)$$

If eq. (4.1) is a good approximation then using estimators for its parameters β yields an accurate predictor $\hat{y}$. This predictor can be checked against the outcome of an actual simulation run, y. More precisely, let x_0 denote the column vector of prespecified values of the independent variables in a new simulation run. (This run was not used in the estimator $\hat{\beta}$, i.e., x_0 is not included in X.) Hence the expected value of the simulation output is predicted by eq. (4.1):

$$\hat{y}_0 = x_0' \cdot \hat{\beta} \qquad (4.2)$$

Using eq. (2.2) yields[10]

$$\text{var } (\hat{y}_0) = \underset{\sim}{x}_0' \cdot \Omega_{\hat{\beta}} \cdot \underset{\sim}{x}_0 \qquad (4.3)$$

where $\Omega_{\hat{\beta}}$ was given in eq. (2.4). The simulation program reads $\underset{\sim}{x}_0$ and yields the output y_0 with its estimated variance s_0^2, based on d_0 degrees of freedom. The significance can be tested through a Student t-statistic:

$$t_d = \frac{\hat{y}_0 - y_0}{[\hat{\text{var}} (\hat{y}_0) + \hat{\text{var}} (y_0)]^{1/2}} \qquad (4.4)$$

where d (the d.f. of t) may be set to d_0.[11]

If GLS with estimator $\hat{\Omega}$ is used then the asymptotic variance of $\hat{\underset{\sim}{\beta}}$ might be used in eq. (4.3).

If the constant variance assumption held, then an F-test for lack-of-fit would be possible. This test compares the estimators s_i^2 to the Mean Squared Residuals of eq. (2.6). Apart from its restrictive assumptions, its power (inverse of β-error) is low, if its d.f. are small, e.g. for $\alpha = 0.05$ $F_{12,\infty}^{\alpha} = 1.75$ but $F_{3,\infty}^{\alpha} = 2.60$. Note that some authors claim that the F-test is insensitive to heterogeneity of variance and to nonnormality, whereas other authors hold different opinions.[12]

If the covariance matrix $\underset{\sim}{\Omega}$ were known, then a similar test for lack-of-fit could be applied in GLS.[13] Some more tests are known but they seem less appropriate.[14]

After the validation run is accepted, it can be added to $\underset{\sim}{X}$ and $\underset{\sim}{y}$ so that $\underset{\sim}{\beta}$ can be estimated more precisely. It seems wise to have $\underset{\sim}{x}_0$ correspond with the "center" of the design (i.e., to have the quantitative factors satisfy $x = 0$) in order to test quadratic effects. Some validation runs should correspond with

x-values occurring in practice, as the use of experimental designs
to specify $\underset{\sim}{X}$ means that the x-values correspond to reasonable,
extreme conditions rather than "common" conditions.[15]

5. SIMULTANEOUS TESTS

In regression analysis a number of tests are made: the estimated
regression model is checked against one or more validation runs,
and individual parameters β are tested. These multiple tests
raise the problem of experimentwise error rates.

In the case study reported on in section 6 ten extra runs
are available to test the adequacy of the regression (meta) model. By
definition the α-error implies that

$$P (t_d \geq t_d^{\alpha} \mid H_0) = \alpha \qquad (5.1)$$

Hence even if the null-hypothesis of an adequate model holds, ten valida-
tion runs are expected to result in one significant t-value if a
traditional α of 10% is used. The simplest solution is to replace
α in eq. (5.1) by α/n where n denotes the number of tests, i.e.
n = 10. Instead of this simple "Bonferroni" approach more complicated
"multiple comparison procedures" are available; see Kleijnen (1975,
pp. 525-597). Note that protection of the α-error (type I error)
increases the β-error (type II error), i.e. it becomes more
difficult to detect an incorrect model specification. Therefore
the experimentwise error rate is usually fixed at a high value
such as 20%.

Next consider the evaluation of separate components of the
model. As an illustration assume that the model incorporates k

factors:

$$y_i = \beta_0 + \sum_{j=1}^{k} \beta_j x_{ij} + e_i \tag{5.2}$$

Then the paramters β_j can be tested through the t-test of eq.
(2.7). Each factor is considered individually, i.e., the experiment's
interpretation does not hinge on the joint results of the tests.
Therefore I propose to stick to the familiar α-rates of, say,
10%: "per comparison" error rate. Remember that in the validation
phase the model is rejected if any validation run yields a significant
t-value: experimentwise error.

Consider another example, in which only two factors are
studied but a more complicated model is postulated:

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_{12} x_{i1} x_{i2} +$$
$$+ \beta_{11} x_{i1}^2 + \beta_{22} x_{i2}^2 + e_i \tag{5.3}$$

Suppose that the t-test of eq. (2.7) shows that all $\hat{\beta}$'s are
significant except for $\hat{\beta}_{11}$. Remember, however, that $\hat{\beta}_{11}$ is an
unbiased estimator of β_{11}; if the assumptions of eq. (1.1) hold, then
$\hat{\beta}_{11}$ is even a BLUE. Strong reasons may exist to formulate a null-hypo-
thesis. For instance, the parsimonous character of scientific models
requires that instead of postulating that "everything depends on
everything else", the observation y be explained by as few factors
as possible: $H_0^{(1)}: \beta_i = 0$ ($i = 1, \ldots, k$). Eq. (5.3),
however, postulates that y is a quadratic polynomial in x_1 and
x_2. Hence I propose to maintain the small, but non-zero, $\hat{\beta}_{11}$-
value.

A different question may arise: can eq. (5.3) be replaced by a __simpler__ model, namely a first degree polynomial in x_1 and x_2? This question can be answered in different ways:

(1) Formulate the __composite__ hypothesis

$$H_0 : \beta_{12} = 0 \wedge \beta_{11} = 0 \wedge \beta_{22} = 0 \qquad (5.4)$$

where $\wedge$ denotes the logical operator "and". The experimentwise error is controlled if a common variance is assumed and the appropriate ANOVA F-test is used.[16]

(2) The hypothesis of eq. (5.4) can also be tested applying the individual t-tests of eq. (2.7) with α replaced by $\alpha/3$: Bonferroni-approach.

(3) A cruder approach estimates the first order polynomial

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + e_i \qquad (5.5)$$

and validates this model with runs not used in estimating eq. (5.5); see eqs. (4.4) and (5.1). Alternative (3) is cruder: if the simpler model of eq. (5.5) is rejected, it is unknown whether this rejection is caused by a large value for β_{12}, β_{11} or β_{22}. See also Kleijnen (1975, pps. 358-367).

6. AN APPLICATION

This section summarizes a case study presented in detail in Kleijnen et al. (1979) and Van den Burg et al. (1977). The present summary emphasizes the statistical techniques applied in that study. Moreover, both previous publications are based on a Monte Carlo experiment that contained a programming bug. The correct results are given in Appendix 1.

Europe Container Terminus (ECT) in the Rotterdam harbour
provides facilities for handling en storing containers. A simulation
model represents storage capacity w as a function of yearly throughput
(production). A fixed production size can be realized by many small
ships or by a few big ships; hence define the mean ship-size x_1
and the arrival rate x_2. Four more factors are investigated: x_3
through x_4, e.g. x_4 denotes the shape (not the mean) of the ship-size
distribution.[17] Every eight simulated hours the simulation gives
a snapshot of the storage size. From this time series w_t (t = 1,
..., T) a frequency diagram is formed. The frequency diagram yields
an average and a few selected quantiles such as the 90% quantile. Fig.
1 is a simplified flow chart of the simulation model. The present
summary concentrates on the average storage capacity y (or $\bar{w} = \Sigma w_t/T$
in the above symbols). The other outputs such as the 90% quantile
$(\hat{w}_{.90})$ are analyzed similarly, although more sophisticated multi-
variate analysis would be better.

The complicated simulation model of Fig. 1 defines a function
f:

$$y = f (x_1, \ldots, x_6, \underset{\sim}{r}) \qquad\qquad (6.1)$$

where $\underset{\sim}{r}$ denotes the random number vector. The complicated function
f is approximated (in the area of experimentation) by a regression
model linear in its parameter β but not necessarily linear in the
variables x. Preliminary studies suggested that the response y
reacts non-linearly to the interarrival time but linearly to the
interarrival rate, i.e. a simple transformation 1/x simplifies the

SHIPGENERATOR
- GENERATE A SHIP
- ATTACH PROPERTIES
- ACTIVATE SHIP
- HOLD

CRANEPLANNER
- PASSIVATE
- ASSIGN CRANES
- CALCULATE FOR ALL SHIPS IN OPERATION END OF OPER
- REACTIVATE SHIPS AT THESE MOMENTS (SOMETIMES OVERRULING AN EVENTTIME)

INSPECTOR
- CALCULATE STATUS OF THE TERMINAL USING QUEUES BA AND AS
- IF ASKED, PRINT STATUS
- PUT DATA IN HISTOGRAMS
- HOLD 8 HOURS

SHIP
- JOIN QUEUE "BEFORE ARRIVAL"
- HOLD 10 DAYS
- LEAVE QUEUE "BEFORE ARRIVAL"
- ROOM FOR SHIP AVAILABLE ?   YES / NO
- IF NOT IN IT, JOIN QUEUE "WAITINGROOM"
- PASSIVATE
- OCCUPY BERTH
- HOLD 1/2 HOUR
- REACTIVATE CRANEPLANNER
- PASSIVATE
- REACTIVATE CRANEPLANNER
- HOLD 1 HOUR
- LEAVE BERTH
- REACTIVATE SHIPS IN WAITINGROOM, IF ANY
- JOIN QUEUE "AFTER SAILING"
- HOLD 10 DAYS
- LEAVE QUEUE "AFTER SAILING"
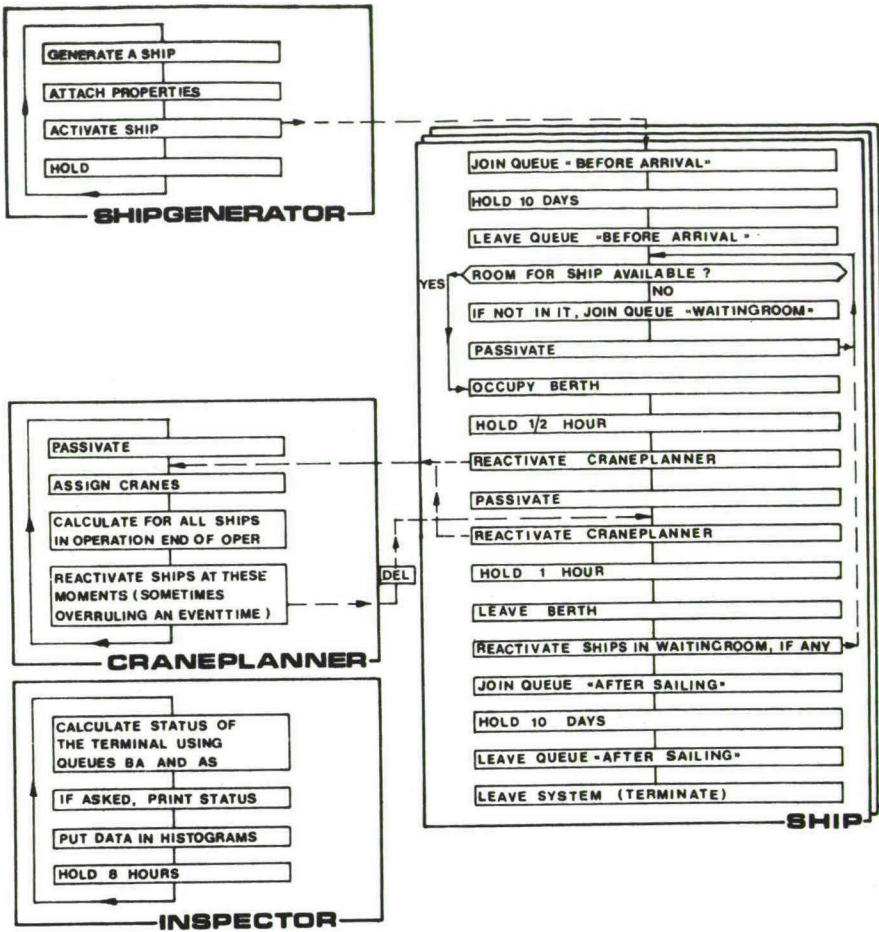- LEAVE SYSTEM (TERMINATE)

DEL

Fig. 1. The simulation model.

model. Quadratic effects (of the quantitative factors x_1 through x_3) are assumed to be zero. Interaction effects between factor 2 and the other factors are suspected to be important: introduce β_{12}, β_{23}, β_{24}, β_{25}, and β_{26}. Moreover, β_{13} may be important. So $\underset{\sim}{\beta}$ comprises one overall mean β_0, six main effects β_1 through β_6, and six interactions, together q = 13 parameters. The selection of an appropriate $\underset{\sim}{X}$ is the domain of experimental design theory; see Kleijnen (1975) and (1979). Application of this theory results in a 16 x 13 X-matrix with all 13 columns orthogonal.[18] So 3 degrees of freedom remain for a possible F-test for lack-of-fit (section 4). However, instead of this F-test the t-test of eq. (4.4) can be applied to ten extra runs executed besides the above 16 x 13 X-matrix. Van den Burg et al. (1977, pp. 57-68) give ten tables containing a great many data. Here only a few tables are presented.

The Monte Carlo experiment suggests that in Table 1 the standard errors for $\hat{\underset{\sim}{\beta}}$ and hence the corresponding t-values may be based on the asymptotic formula eq. (3.1). Table 2 shows that the OLS regression model need not be rejected, since the maximum of the ten t-statistics is 1.67 whereas the significance level is 2.33 for $\alpha = 0.20/10$ (experimentwise error of 20%). If for GLS the asymptotic eq. (3.1) is used, then the validation runs need not be rejected either (not shown in tables). If the F-test for lack-of-fit were used (assuming constant variances) then the OLS model would be rejected: $F_{3,128} = 3.50$ is significant at $\alpha = 0.05$. The estimated variances s_i^2, however, vary drastically, namely between 64 and 93,228 (in the first sixteen runs; in the next ten runs s_i^2 varies between 152 and 22,102; see column "vâr(y)" in Table 2).

Table 1

OLS and GLS estimators of regression parameters $\underset{\sim}{\beta}$ based on $2^{8-4} = 16$ runs

a)	OLS $\hat{\beta}$ b)	s c)	t d)	GLS $\tilde{\underset{\sim}{\beta}}$ e)	s f)	t d)
0	-1.420	112.483	-0.013	27.434	30.341	0.904
1	-0.769	15.960	-0.048	-6.656	3.845	-1.575
2	13.440	38.420	0.350	28.566	22.639	1.262
3	-11.508	24.814	-0.479	-17.108	8.849	-1.933
4	3.500	16.042	-0.218	9.267	14.750	0.628
5	-1.375	16.042	-0.086	4.138	14.851	0.279
6	140.918	96.256	1.464	151.932	67.672	2.245*
1,2	15.391	3.192	4.621**	14.644	2.089	7.009**
1,3	0.046	3.331	0.014	1.152	0.896	1.285
2,3	281.098	6.662	42.196**	280.352	5.931	47.268**
2,4	21.250	13.323	1.595	10.729	11.858	0.905
2,5	11.875	13.323	0.891	6.560	11.922	0.550
2,6	-49.483	79.939	-0.619	-139.107	50.129	-2.775

a) 0 refers to β_o; ...; 2,6 refers to β_{26}

b) see eq. (1.2)

c) standard error: square root of main diagonal element in eq. (2.4)

d) Student t- statistic; eq. (2.7)

e) see eq. (1.3) substituting $\hat{\Omega}$

f) asymptotic eq. (3.1)

** significant at any $\alpha > 0.00005$

* significant at $\alpha = 0.025$

Table 2

Model Validation (OLS)

y	$\hat{y}$ [a]	$y - \hat{y}$	$\hat{var}(y)$ [b]	$\hat{var}(\hat{y})$ [c]	t [d]
8,332	8,715	-383	22,102	30,494	-1.67
3,002	2,919	83	1,156	5,092	1.05
729	743	-14	544	964	-0.36
1,725	1,774	-49	625	1,142	-1.16
1,893	1,814	79	4,444	1,205	1.05
685	684	1	107	847	0.03
2,977	3,058	-81	4,761	8,308	-0.71
8,469	8,415	54	10,885	20,808	0.30
608	595	13	152	920	0.40
1,674	1,624	50	514	1,138	1.23

a)   eq. (4.2)

b)   simulation run divided into 9 subruns

c)   eq. (4.3)

d)   eq. (4.4)

After accepting the regression model, the ten validation runs are included in $\underset{\sim}{X}$ and $\underset{\sim}{y}$, and $\underset{\sim}{\beta}$ is reestimated. The effects $\hat{\beta}_{23}$ and $\hat{\beta}_{12}$ remain very significant, namely $t = 49$ and 5.5 respectively. Using GLS their significance further increases: $t = 64$ and 7.9.

Whereas $\underset{\sim}{X}$ comprises standardized variables ($x = +1$ or $x = -1$) the actual design and regression model comprise "user" variables z_1, e.g., z_1 is either 200 or 1,000. The user variables have as significant parameters $\hat{\beta}_{23}$ and $\hat{\beta}_{12}$, whereas the standardized variables would have significant parameters $\hat{\gamma}_0$, $\hat{\gamma}_2$, $\hat{\gamma}_3$, $\hat{\gamma}_{23}$, $\hat{\gamma}_1$ and $\hat{\gamma}_{12}$ (in order of decreasing significance, using γ to denote the parameters of the standardized variables x).[19]

So some parameter estimates $\hat{\underset{\sim}{\beta}}$ are found insignificant (after validating the first sixteen runs using ten extra runs, and then reestimating $\underset{\sim}{\beta}$ from all twenty-six runs). Next these insignificant parameters are set to zero, and the remaining $\underset{\sim}{\beta}$ (i.e. β_{23} and β_{12}) are again reestimated.[20]

In general, it is recommended to examine the residuals $y - \hat{y}$ to see whether they satisfy the classical assumptions of eq. (1.1); see Draper & Smith (1966, pp. 86-94). Upon studying the responses, and especially the residuals, applying just "common sense", certain patterns emerged. These patterns suggest the importance of interactions until then ignored, namely β_{14} and β_{15} (see again note 18). Fortunately, incorporating these two new effects into $\underset{\sim}{X}$ left $\underset{\sim}{X}$ non-singular (see also next section). The resulting $\hat{\underset{\sim}{\beta}}$ still contains as significant parameters $\hat{\beta}_{23}$ and $\hat{\beta}_{12}$ only.

Instead of <u>backwards elimination</u> of insignificant parameters, one might proceed from the other direction. In <u>stepwise regression</u> one new variabele is introduced in each step, namely the (remaining) variabele x showing maximum correlation with the dependent variable y. The qualitative results are similar to backwards elimination; first β_{23} is introduced, then β_{12} (and next β_{14} etc.)

The above procedure is summarized in Fig. 2. The discussion should make it obvious that the procedure cannot be used mechanically. The selection of variables in regression models is discussed from a statistical viewpoint by Hocking (1976); see also Enslein et al. (1977) and Thompson (1978). However, the regression model's specification involves more than a statistical bag of tricks. Model specification requires intuition, and prior knowledge based on relevant theories and empirical data. In the present case study the most significant parameter β_{12} was also the one parameter suggested by a simplified analytical model.

7. ALTERNATIVES TO OLS AND GLS

Both OLS and GLS use as criterion minimization of squared residuals: least squares. Simulation practitioners tend to focus on <u>relative</u> residuals: $|y - \hat{y}|/y$. This criterion leads to a linear programming problem: see Narula & Wellington (1977). Unfortunately, the properties of the resulting estimators are unknown, whereas for OLS or GLS the estimators are known to be BLUE, and a battery of statistical tests is available.

The choice of the criterion also affects the sensitivity of the resulting estimates to outliers, i.e., wild observations on y

Figure 2: Summary of regression procedure
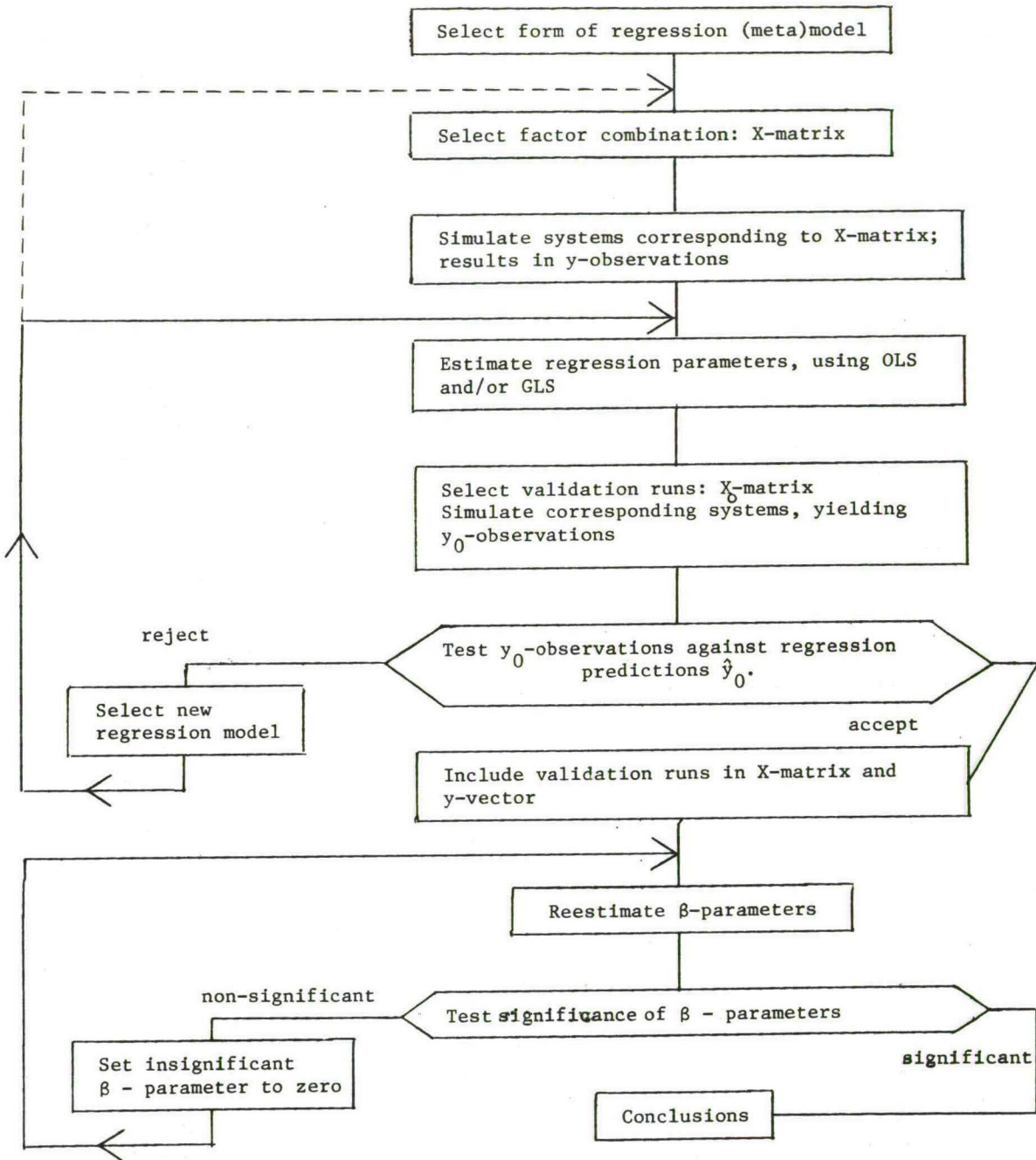
or x. Denby & Larsen (1977) give a survey of robust regression estimators and they present Monte Carlo results for a number of these estimators. A special issue on robustness including a number of articles on robust regression analysis, was edited by Hogg (1977).

If the X-matrix is ill-conditioned, ridge estimation may be of interest, i.e., the estimators of β are no longer unbiased; however, this bias may be outweighted by a decrease in variance through a proper choice of the ridge algorithm parameters; see Hocking (1976). In simulation X can be made orthogonal, but introducing unexpected parameters (such as β_{14} and β_{15} in the preceding section) can make X perfectly or nearly singular.[21]

Dempster et al. (1977) performed an extensive simulation experiment (160 data sets) examining 57 different regression estimators!

Instead of selecting an appropriate estimation algorithm, a matrix of independent variables X can be selected so that the sensitivity of the estimates towards outliers is minimized; see Box & Draper (1975).

One more alternative is provided by the Bayesian decision-theoretic model: prior probabilities on parameters like β are postulated (Bayes approach), together with loss functions like $\Sigma w_i (\beta_j - \hat{\beta}_j)^2$. Instead of fixing the α-values, the expected "aposterior" (after taking the sample) loss is minimized, or the maximum loss is minimized; see Dempster et al. (1977).

8. CONCLUSION

To mitigate the ad-hoc character of simulation, regression analysis can be used so that a metamodel results. The metamodel serves to interpret the simulation results.

The regression analysis can use OLS or GLS. When applying OLS the experimenter should check for non-constant variances σ_i^2 (estimated from the individual simulation runs). Heterogeneity of variances changes the formula for $\hat{\Omega}_\beta$ (the covariance matrix of the estimated parameters $\hat{\beta}$) and hence the corresponding t-test for significance. A Monte Carlo experiment suggests that GLS with estimated Ω (covariance matrix of the observations) results in a covariance matrix for the β-estimators that can be approximated accurately by the asumptotic formula, eq. (3.1).

The regression metamodel's validity can be tested statistically applying a t-test. Multiple validation runs raise the issue of experimentwise error rates. This complication may be solved through the Bonferroni inequality.

The model form and the values specified in null-hypotheses have to come from non-statistical sources such as engineering and management science. Subjective elements remain in the selection of the α-values and in the evaluation of the statistical technique's sensitivity to assumptions like normality and constant variance.

APPENDIX 1: MONTE CARLO EXPERIMENT WITH OLS AND GLS

The Monte Carlo experiment computes the observations y from the "true" model

$$\underset{\sim}{y} = \underset{\sim}{X} . \underset{\sim}{\beta} + \underset{\sim}{e} \tag{A1.1}$$

where $\underset{\sim}{X}$ and $\underset{\sim}{\beta}$ are known inputs and $\underset{\sim}{e}$ has a known distribution

$$\underset{\sim}{e} \sim N(\underset{\sim}{0}, \underset{\sim}{\Omega} = \underset{\sim}{D}) \tag{A1.2}$$

$\underset{\sim}{D}$ being a diagonal matrix with elements σ_i^2. Next - as in practice - only $\underset{\sim}{y}$ and $\underset{\sim}{X}$ (together with an estimate $\hat{\Omega}$) are made available to the OLS and GLS algorithms to compute the estimates $\hat{\underset{\sim}{\beta}}$ and $\hat{\hat{\underset{\sim}{\beta}}}$, respectively. Finally, the latter estimates are compared to the true values $\underset{\sim}{\beta}$.

In simulation $\underset{\sim}{\Omega}$ is estimated from the m_i individual subruns within a simulation run:

$$s_i^2 = \sum_{j=1}^{m_i} (y_{ij} - \bar{y}_i)^2 / (m_i - 1) \tag{A1.3}$$
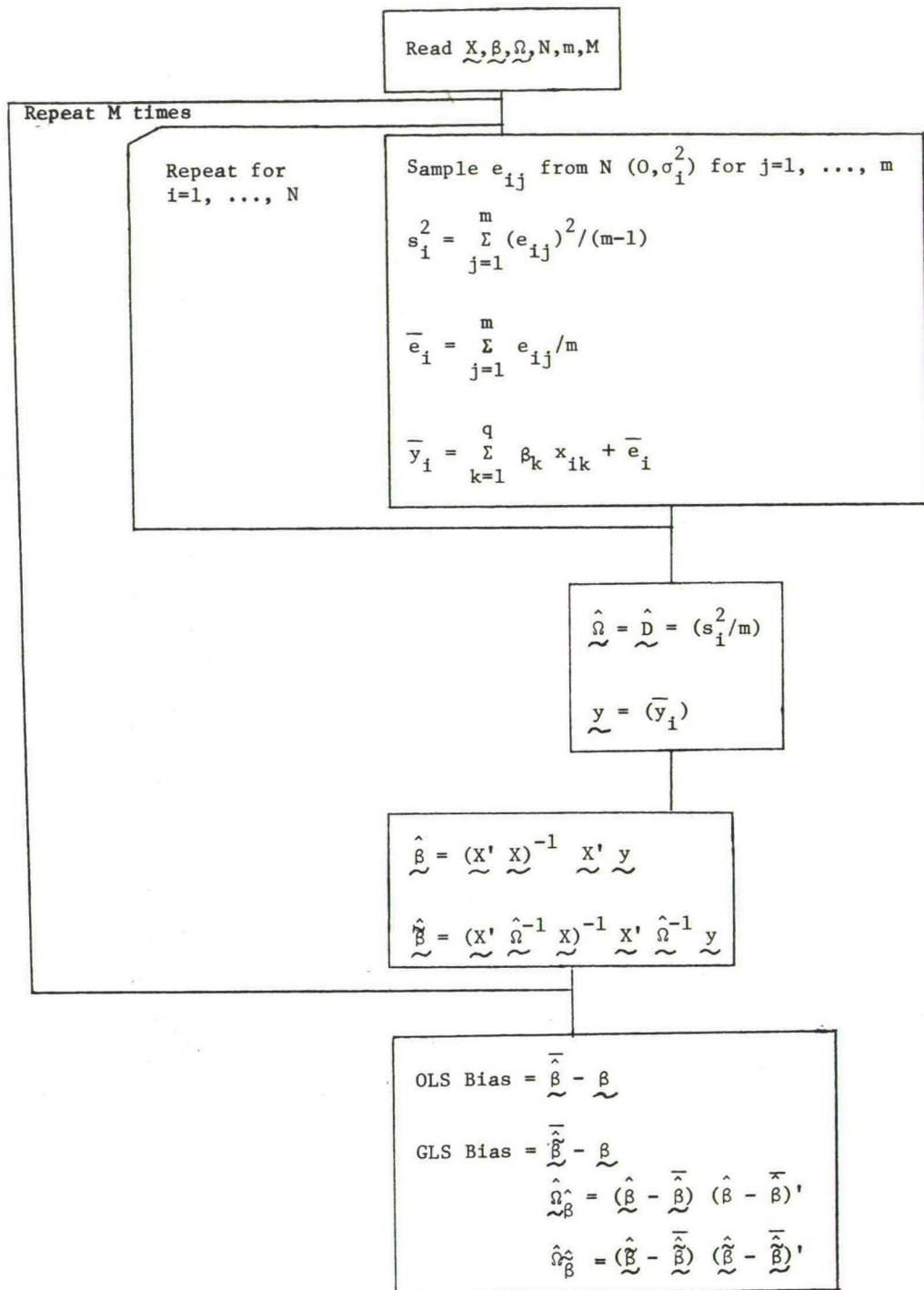
Hence s_i^2 is a χ^2-variable with $d_i = m_i - 1$ degrees of freedom (d.f.) and s_i^2 can be sampled using the simpler formula

$$s_i^2 = \sum_{j=1}^{m_i} e_{ij}^2 / (m_i - 1) \tag{A1.4}$$

The OLS and GLS estimators use not the subrun observations y_{ij} but only the total-run observations $\bar{y}_i$. Consequently

$$\sigma_i^2 = var (\bar{y}_i) = var (y_{ij}) / m \tag{A1.5}$$

Figure 3: Summary of Monte Carlo experiment

Read $\underset{\sim}{X}, \underset{\sim}{\beta}, \underset{\sim}{\Omega}, N, m, M$

Repeat M times

Repeat for
$i=1, \ldots, N$

Sample e_{ij} from $N(0, \sigma_i^2)$ for $j=1, \ldots, m$

$$s_i^2 = \sum_{j=1}^{m} (e_{ij})^2 / (m-1)$$

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

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

$$\underset{\sim}{\hat{\Omega}} = \underset{\sim}{\hat{D}} = (s_i^2 / m)$$

$$\underset{\sim}{y} = (\bar{y}_i)$$

$$\underset{\sim}{\hat{\beta}} = (\underset{\sim}{X}' \underset{\sim}{X})^{-1} \underset{\sim}{X}' \underset{\sim}{y}$$

$$\underset{\sim}{\hat{\tilde{\beta}}} = (\underset{\sim}{X}' \underset{\sim}{\hat{\Omega}}^{-1} \underset{\sim}{X})^{-1} \underset{\sim}{X}' \underset{\sim}{\hat{\Omega}}^{-1} \underset{\sim}{y}$$

OLS Bias $= \underset{\sim}{\bar{\hat{\beta}}} - \underset{\sim}{\beta}$

GLS Bias $= \underset{\sim}{\bar{\hat{\tilde{\beta}}}} - \underset{\sim}{\beta}$

$$\underset{\sim}{\hat{\Omega}}_{\hat{\beta}} = (\underset{\sim}{\hat{\beta}} - \underset{\sim}{\bar{\hat{\beta}}})(\underset{\sim}{\hat{\beta}} - \underset{\sim}{\bar{\hat{\beta}}})'$$

$$\underset{\sim}{\hat{\Omega}}_{\hat{\tilde{\beta}}} = (\underset{\sim}{\hat{\tilde{\beta}}} - \underset{\sim}{\bar{\hat{\tilde{\beta}}}})(\underset{\sim}{\hat{\tilde{\beta}}} - \underset{\sim}{\bar{\hat{\tilde{\beta}}}})'$$

The Monte Carlo experiment is summarized in Fig. 3.

Fig. 3 shows that a number of values has to be fixed in the Monte Carlo experiment. As true values for $\underset{\sim}{\beta}$ are selected the estimates from the E.C.T. study; see the column labelled "$\hat{\beta}$" in Table 1. $\underset{\sim}{X}$ is the corresponding 16 x 13 orthogonal matrix (see again note 18); hence N = 16. Further m = 9 and e_{ij}^2 is sampled from $N(0, \sigma_i^2)$ with σ_i^2 equal to m times the square of "s" in Table 1, see eq. (A.1.4). M is arbitrarily set to 250, and turns out to yield significant differences between OLS and GLS. In a second experiment the ten validation runs are included, and hence N = 26. The "true" $\underset{\sim}{\beta}$ corresponds with the $\hat{\underset{\sim}{\beta}}$ estimated from the augmented $\underset{\sim}{X}$ and $\underset{\sim}{y}$ matrices in the E.C.T. study. Normal variates are sampled using the familiar Box-Muller trigonometric transformation.

The Monte Carlo results can be summarized as follows.

(1) OLS yields unbiased estimators of $\underset{\sim}{\beta}$, even if $\underset{\sim}{\Omega} \neq \sigma^2 \underset{\sim}{I}$. This property can be proved analytically; see note 4. The Monte Carlo experiment double-checks this property, computing per effect $\underset{\sim}{q}$:

$$z_q = \frac{\overline{\tilde{\beta}_k} - \beta_k}{[var\ (\hat{\beta}_k)/M]^{1/2}} \qquad (q=1,\ \ldots,13) \qquad (A1.6)$$

which should be distributed as a standard normal variable N(0, 1). Note that var($\hat{\beta}_q$) follows from eq. (2.4), in which $\underset{\sim}{\Omega}$ is known. The maximum z turns out to be 1.54, whereas the critical z-value is 2.16 for an experimentwise error rate of 20% (so that the individual α becomes 0.20/13 = 1.54%)

Table 3: Monte Carlo results

β [a]	Theoretical σ^2		Monte Carlo [d] deviation	
	OLS [b]	GLS [c]	OLS(%)	GLS(%)
1	710.042	297.212	-14.17	- 1.77
2	710.042	568.936	-18.11	- 4.98
3	710.042	566.836	-14.87	- 8.53
4	710.042	562.356	12.27	11.18
5	710.042	571.266	22.02	20.39
6	710.042	282.799	7.75	2.44
1,2	710.042	279.342	- 6.34	9.73
1,3	710.042	51.392	- 4.53	-11.69
2,3	710.042	562.867	-17.78	- 8.64
2,4	710.042	562.425	12.61	15.12
2,5	710.042	568.574	20.64	19.95
2,6	710.042	279.214	6.19	2.64
0	710.042	568.940	-19.23	- 9.47

a) 1 corresponds with β_1, ς.. 2,6

   with β_{26}, and 0 with β_0.

b) main diagonal of eq. (2.4)

c) main diagonal of eq. (3.1)

d) $\left\{(\hat{\sigma}^2 - \sigma^2)/\sigma^2\right\}$ x 100 where $\hat{\sigma}^2$ denotes the Monte Carlo

   estimate; see eq. (A1.7)

(2) GLS with estimated $\underset{\sim}{\Omega}$ remains unbiased. In eq. (A1.6) $\text{var}(\hat{\beta}_k)$ is replaced by the estimator

$$\widehat{\text{var}}(\hat{\beta}_q) = \sum_{g=1}^{M} (\hat{\beta}_{qg} - \bar{\hat{\beta}}_q)^2 \, /(M-1) \qquad (A1.7)$$

and z is replaced by t_{M-1}, the Student t-statistic with M-1 d.f. . However, since M=250 the t-distribution can be approximated by the z-distribution. The extreme t-value is -1.99: non-significant. Since the OLS and GLS algorithms use the same data, a positive (negative) deviation $\hat{\beta} - \beta$ tends to be accompanied by a positive (negative) deviation $\hat{\hat{\beta}} - \beta$: positive correlation between $\hat{\beta}$ and $\hat{\hat{\beta}}$.

(3) The OLS estimators $\hat{\beta}$ have a known covariance matrix $\underset{\sim}{\Omega}_{\hat{\beta}}$; see eq. (1.3). The theoretical result agrees with the estimator $\hat{\underset{\sim}{\Omega}}_{\hat{\beta}}$ computed from M=250 replications: Test each diagonal element s^2 of $\hat{\underset{\sim}{\Omega}}_{\hat{\beta}}$ versus the corresponding element σ^2 in $\underset{\sim}{\Omega}_{\hat{\beta}}$ via the χ^2-statistic:

$$\chi^2 = \frac{s^2}{\sigma^2} \cdot (df) \qquad (A1.8)$$

where df = M-1 = 249. The maximum s^2/σ^2 occurs for β_5 and equals 1.22; see Table 3. The value 1.22 just remains insignificant with an experimentwise α of, say, 20%.[22] The covariances between two OLS estimators are positive (not displayed). This positive correlation is explained by the specific values of $\underset{\sim}{X}$ and $\underset{\sim}{\Omega}$ in this Monte Carlo experiment; see note 6.

(4) For GLS using estimator $\hat{\underset{\sim}{\Omega}}$ only an asymptotically valid covariance formula is known; see eq. (3.1). The Monte Carlo experiment estimates the small-sample (m=9) variances using eq. (A1.7). These estimates based on M=250 replications, are compared to the asymptotic variances. As Table 3 shows the maximum deviation occurs again for β_5, and is

20%, i.e. the maximum deviation is smaller than that for OLS. Not only is the maximum error smaller, but as Table 3 shows for 10 out of 13 parameters the deviation between estimated and theoretical variances is smaller for GLS than for OLS. Note that the GLS estimators are sometimes negatively correlated (not displayed).

(5) Comparison of the estimated covariance matrices of the OLS and GLS estimators, shows that the GLS estimators of $\underset{\sim}{\beta}$ have smaller standard errors than the OLS estimators have. Moreover, all estimated covariances are also smaller in GLS than in OLS. In shorthand notation:

$$\hat{\underset{\sim}{\Omega}}_{\hat{\underset{\sim}{\beta}}} < \hat{\underset{\sim}{\Omega}}_{\hat{\underset{\sim}{\beta}}} \qquad (A1.9)$$

where the "smaller than" relation holds for all elements of the two 13 x 13 matrices. This inequality might be tested statistically, using a result derived by Box; see Morrison (1967, pp. 152-153).[23]

(6) The whole experiment is repeated once more. The 16 x 13 X-matrix is augmented with ten extra runs. New values for all thirteen β's are used; $\underset{\sim}{\Omega}$ is augmented to a 26 x 26 matrix. The results are as follows. For GLS the maximum s^2/σ^2 is now 1.218 which is insignificant; see the discussion below eq. (A1.8). It is interesting to note that for all 13 parameters the asymptotic formula yields variances smaller than the Monte Carlo estimates, though insignificantly smaller according to the x^2 test of eq. (A1.8). This suggests that the asymptotic formula slightly underestimates the true variance. Hence estimated β-parameters found to be significant may actually be insignificant (insignificant parameters remain insignificant). Again the relation (A1.9) is found.

NOTES

1. The estimated variances s_i^2 ($i=1, \ldots, N$) could be subjected to a statistical test with $H_0 : \sigma_1^2 = \ldots = \sigma_N^2 \ (= \sigma^2)$. Finding a robust test (i.e., a test not very sensitive to its underlying assumptions such as normality) may be difficult; see Scheffé (1964). Even if the test is robust, it may still have little power, i.e., the test may not be able to detect deviations from the null-hypothesis.

2. ANOVA with fixed effects is a special case of linear regression analysis; see Kleijnen (1975, p. 301).

3. A referee pointed out that the Biomed computer package allows for Generalized Least Squares; see [3, p. 453] and eq. (1.3). Nevertheless my experience has been that in the simulation field practitioners do not use such options. The purpose of this tutorial is exactly to bring these options to their attention.

4. $\&(\hat{\underset{\sim}{\beta}}) = \&(\underset{\sim}{W} \underset{\sim}{y}) = \underset{\sim}{W} \&(\underset{\sim}{y}) = (\underset{\sim}{X}' \ \underset{\sim}{X})^{-1} \underset{\sim}{X}' \ (\underset{\sim}{X} \ \underset{\sim}{\beta}) = \underset{\sim}{\beta}$

5. If renewal analysis is applied to estimate σ_i^2, then use $d_i = \infty$, because renewal analysis is based on asymptotic formulas. For the estimation of standard errors of autocorrelated simulation responses see Kleijnen (1979).

6. $\underset{\sim}{X}$ is orthogonal means $\underset{\sim}{X}' \ \underset{\sim}{X} = N \ \underset{\sim}{I}$. Hence eq. (1.2) reduces to

$$\hat{\beta}_j = \sum_{i=1}^{N} x_{ij} \ y_i / N \qquad (j=1, \ldots, q)$$

and eq. (2.4) becomes

$$\underset{\sim}{\Omega}\ \hat{\beta} = \underset{\sim}{X}' \ \underset{\sim,\sim}{D \ X} \ / \ N^2$$

From the main diagonal elements of $\Omega_{\hat{\beta}}$, or directly from the pre-ceding equation for $\hat{\beta}_j$, it follows that

$$\text{var} \ (\hat{\beta}_j) = \overset{N}{\underset{i=1}{\Sigma}} \ x^2_{ij} \ \ \sigma^2/N^2$$

Hence, even though $\Sigma x_{ij} \ x_{ij'} = 0$ for $j \neq j'$, the covariances do not reduce to zero if $\sigma^2_i \neq \sigma^2$.

7. If common random numbers or antithetic numbers were used, then $\underset{\sim}{\Omega}$ would no longer be a diagonal matrix; Kleijnen (1975) and (1979).

8. Under the standard assumptions of eq. (1.1) any textbook on regression analysis shows that the d.f. of t equal those of $\hat{\text{var}} \ (\hat{\beta})$. Because $\hat{\text{var}} \ (\hat{\beta})$ follows from eqs. (2.5) and (2.6), the d.f. of $\hat{\text{var}} \ (\hat{\beta})$ equal (N-q). In simulation assuming constant variances $\sigma^2_i = \sigma^2$ means that the d.f. of $\hat{\text{var}} \ (\hat{\beta})$ become Σd_i, pooling the d.f. of the individual estimators s^2_i. If the s^2_i have different expectations σ^2_i, then I conjecture that a test is possible using as d.f. for t the <u>minimum</u> of the d_i -values; see Scheffé (1970, p. 1502) for the "conservative" character of such an approach, i.e., the actual type-I error may be smaller than the nominal α-value. Observe further that the numerator and denominator of the t-statistic defined by eq. (2.7) remain independent: $\hat{\beta}$ is a function of the y_i whereas $\hat{\text{var}}(\hat{\beta})$ is a function of the s^2_i, and y_i and s^2_i are independent, provided the subrun averages as assumed normally distributed.

9. See also Kleijnen (1975, pp. 719-720), and Schmidt (1976, pp. 65-72).

10. Observe that Hocking (1976) introduces the predicted individual
observation y + e with variance equal to var (y) + var (e); see
his equations (2.13) and 4.1).

11. Originally the t-test was derived for comparing a single sample
average with its hypothesized value, or for comparing two indepen-
dent sample averages from two populations with each other provided
the two populations have a common variance. I conjecture that a
conservative test may be based on eq. (4.4) with the d.f. equal to
the minimum of the d.f. of $\widehat{var}(\hat{y}_0)$ and $\widehat{var}(y_0)$; see footnote 8 and
Kleijnen (1975, pp. 470-472). The minimum d.f. equal d_0, unless
the verification run is extremely long. Note that the t-approxima-
tion suggests making $\hat{y}_0$ and y_0 independent, i.e., excluding $\underset{\sim}{x}_0$
from $\underset{\sim}{X}$. Observe that some non-statistical publications like Keeney
& Raiffa (1976, pp. 280-281) compare not y to $\hat{y}$, but the actual
changes $y_i - y_i$, to the predicted changes $\hat{\hat{y}}_i - \hat{y}_i$. However, this
increases the variance of the relevant statistic.

12. MSR defined in eq. (2.6) has expected value σ^2 if the model of eq.
(4.1) is correct (and the standard assumptions of eq. (1.1) hold);
else $\mathcal{E}(MSR) > \sigma^2$. Simulation run i (or in general, replication of
factor combination i) yields the estimator s_i^2, unbiased even if
eq. (4.1) is not a good (meta) model. Hence $F_{d_1, d_2} = MSR/s_i^2$ with
$d_1 = N-q$ and $d_2 = d_i$. Because of the common variance assumption,
different runs can be pooled: $\sum_1^N s_i^2/N$ with Σd_i d.f. . In simu-
lation the denominator's d.f. are usually high (in the case-study

16 x 8 = 128) but the numerator's d.f. is low, if the number of simulation runs is small compared to the number of regression para-meters β (in the case-study 16 − 13 = 3). In GLS a similar procedure applies; see next footnote. For a related test see Lyons & Proctor (1977). For the F-test's robustness and power see Bishop & Dudewicz (1977, pp. 1, 24-25), Derby & Larsen (1977, p. 250), and Kleijnen (1975, p. 725).

13. The original regression model − see eq. (4.1) − implies

$$V.y = V.X.\beta + V.e$$

Denote the transformed variables by * so that

$$y* = X*.\beta + e*$$

It is possible to select V such that

$$V'.V = \Omega^{-1}$$

Eq. (2.2) yields then

$$\Omega* = V.\Omega.V' = I$$

Hence the transformed variables $y*$ satisfy the classical assumptions with $\sigma^2 = 1$. Therefore there is no need to estimate σ^2 in the denominator of the F-statistic. Consequently the F-statistic (a ratio of two variance estimators) becomes a χ^2-statistic:

$$\chi^2_{n-q} = \sum_{i=1}^{N} (y*_i - \hat{y*_i})^2$$

where $y*$ and $\hat{y*}$ are based on the transformation matrix V which in turn assumes a known matrix Ω. It is unknown whether the χ^2-statistic remains valid if Ω (and hence V) is estimated. See also Schmidt (1976, p. 68), substituting n = T, q = K, $\Omega = \sigma^2 \Omega$ with $\sigma^2 = 1$; it is easy to check that in the classical case where $\Omega = \sigma^2 I$, Schmidt's $\hat{\sigma}^2$ reduces to the familiar expression $\Sigma (y_i - \hat{y}_i)^2/(N-q)$.

14. Hocking (1976, pp. 22-23) proposes to pool all, say, c validation runs: $\sum_{g=1}^{c} (y_g - \hat{y}_g)^2 /c$ and to compare this statistic to MSR. This approach, however, assumes a common variance.

15. A trick to obtain validation runs is to delete one run i from the N old observations, yielding $\underset{\sim}{y}^{(i)}$ and $\underset{\sim}{X}^{(i)}$. Use $\underset{\sim}{y}^{(i)}$ and $\underset{\sim}{X}^{(i)}$ to compute $\hat{\underset{\sim}{\beta}}^{(i)}$. Then $\hat{\underset{\sim}{\beta}}^{(i)}$ can be used to predict y_0. See jackknifing in Kleijnen (1975) and also Hocking (1976, pp. 22-23), Narula & Wellington (1977, p. 188).

16. Pool the sums of squares corresponding with β_{12}, β_{11}, and β_{22} and divide by the sum of the corresponding d.f. Next compare this ratio to an independent estimate of pure error; see Kleijnen (1975, pp. 298, 730).

17. A qualitative factor like distribution shape is represented by the dummy variable x with values +1 (denoting exponential distribution) and -1 (denoting constant distribution).

18. Readers familiar with experimental design, can construct this X-matrix as follows. Use the generators 1 = 56 and 3 = 45. Then the defining relation is: I = 156 = 345 = 1346. Hence the alias pattern, ignoring interactions among three or more factors, is: 1 = 56, 3 = 45, 4 = 35, 5 = 16 = 34, and 6 = 15 (all two-factor interactions in this alias relation were assumed zero).

19. The standardized variable x_2 (arrival rate) is transformed into the user variable z_2 (yearly production), applying $z_2 = x_1 \cdot x_2 \cdot c$ where c is a constant; see Kleijnen et al. (1979, p. 59) and Van den Burg et al. (1977, p. 56-57).

20. What effect will the reestimation of the parameters $\underset{\sim}{\beta}$ have on the lack-of-fit F-test? By definition the least-squares algorithm minimizes the MSR numerator; see eq. (2.6). Imposing the restriction that some parameters are zero, cannot decrease this numerator. In general the numerator increases but this increase will be small, since insignificant $\hat{\beta}$'s are selected. The MSR denominator increases as q decreases. The net result in some experiments turned out to be a lower MSR. Since the denominator of the F-test for lack-of-fit remains the same, a lower F-value resulted. For instance, in the case-study q, the number of parameters, was reduced from seventeen to eight, and this reduction of q decreased the F-statistic from 1.18 to 0.92 (not shown in tables). This F-value was compared to the upper α-point of the F-statistic, with the same degrees of freedom d_2 for the denominator but with a higher d_1 for the numerator. For instance, with $\alpha = 0.05$ corresponds $F_{3,\infty} = 2.60$ but $F_{8,\infty} = 1.94$. So the F-value decreased but it was compared to a lower α-point.

Note that though the expected value of F equals one (under the hypothesis of a correct model), this does not mean that the computed (sampled) F-value cannot be smaller than one. For instance, the F-tables show that $P(F_{3,\infty} < 0.789) = 0.50$. Indeed, in many case studies I found F-values smaller than one.

21. For numerical accuracy when $\underset{\sim}{X}$ is ill-conditioned see Lawson & Hanson (1974) and Miller (1978). For the statistical aspects see Einslein et al. (1977), Hocking (1976), Lawless (1978), Schmidt (1976, pp. 41-53) and Swamy et al. (1978).

22. A two-sided test with $\alpha = 0.20/13 = 1.54\%$ is performed. Hence the upper 99.23 quantile of χ^2_{249} is needed. The χ^2-table shows for d.f. = 250 a value of 1.22 for the 99.0 quantile and 1.25 for the 99.5 quantile.

23. Let H_0: $\underset{\sim}{\Omega}_1 = \underset{\sim}{\Omega}_2$ where $\underset{\sim}{\Omega}_1$ (and $\underset{\sim}{\Omega}_2$) is the 13 x 13 covariance matrix of $\hat{\beta}$ (and $\hat{\tilde{\beta}}$). Then a χ^2-variable results with $1/2\ q\ (q+1)$ d.f., setting $\chi^2 = MC$ with

$$M = (\Sigma n_i)\ \ln|\hat{\underset{\sim}{\Omega}}| - \Sigma n_i\ \ln|\hat{\underset{\sim}{\Omega}}|$$

where $n_i = M-1$ and $\hat{\Omega} = \Sigma n_i \Omega_i /(\Sigma n_i)$, and

$$c^{-1} = 1 - \frac{3(2q^2 + 3q-1)}{6(q + 1)\ 2n}$$

In the Monte Carlo experiment $q = 13$ and $\chi^2_{91} = 641.5$ which has a probability of occurence smaller than 0.1%. Hence H_0 is rejected. Actually GLS and OLS use the same data (same error terms e), and $\hat{\underset{\sim}{\Omega}}_1$ and $\hat{\underset{\sim}{\Omega}}_2$ become dependent. Assume that this dependence reduces the variation in χ^2. Hence the significant χ^2-value of 641.5 becomes even more significant.

REFERENCES

1. BISHOP, T.A. and E.J. DUDEWICZ, HETEROSCEDASTIC ANOVA. Technical
    Report No. 143, Department of Statistics, The Ohio State University,
    Columbus, 1977.

2. BOX, G.E.P. and N.R. DRAPER. Robust designs. BIOMETRIKA, 62, no. 2,
    1975, pp. 347-352.

3. COMPUTER PROGRAMS, University of California Press, 1975.

4. DEMPSTER, A.P., M. SCHATZOFF and N. WERTMUTH, A simulation study of
    alternatives to ordinary least squares. JOURNAL AMERICAN
    STATISTICAL ASSOCIATION, 72, 1977, pp. 77-106.

5. DENBY, L. and W.A. LARSEN, Robust regression estimators compared via
    Monte Carlo. COMMUNICATIONS IN STATISTICS, Part A, 6, no. 4, 1977,
    pp. 335-362.

6. DRAPER, N.R. and H. SMITH, APPLIED REGRESSION ANALYSIS. John Wiley &
    Sons, Inc., New York, 1966.

7. ENSLEIN, K., A. RALSTON and H.S. WILF, STATISTICAL METHODS FOR DIGI-
    TAL COMPUTERS. John Wiley & Sons, Inc., New York, 1977.

8. HOCKING, R.R., The analysis and selection of variables in linear re-
    gression. BIOMETRICS, 32, part 1, March 1976, pp. 1-49.

9. HOGG, R.W., Robustness. Special issue of COMMUNICATION IN STATISTICS-
    THEORY AND METHODS, volume A6, no. 9, 1977, pp. 789-894.

10. KEENEY, R.L. and H. RAIFFA, DECISIONS WITH MULTIPLE OBJECTIVES:
    PREFERENCES AND VALUE TRADEOFFS. John Wiley & Sons, Inc., New York,
    1976.

11. KLEIJNEN, J.P.C., STATISTICAL TECHNIQUES IN SIMULATION. (In two
    volumes). Marcel Dekker, Inc., New York, 1974/1975.

12. KLEIJNEN, J.P.C., The role of statistical methodology in simulation.
    In: SYMPOSIUM ON MODELLING AND SIMULATION METHODOLOGY, edited by
    B. ZEIGLER , North-Holland Publishing Company, Amsterdam, 1979.

13. 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.

14. LAWLESS, J.F., Ridge and related estimation procedures: theory and practice. COMUNICATIONS IN STATISTICS, THEORY AND METHODS, vol. A7, no. 2, 1978, pp. 139-164.

15. LAWSON, C.L. and R.J. HANSON, SOLVING LEAST SQUARES PROBLEMS. Prentice-Hall Book Company, Englewood Cliffs, 1974.

16. LYONS, N.I. and C.H. PROCTOR, A test for regression function adequacy. COMMUNICATIONS IN STATISTICS, Part A, 6, no. 1, 1977, pp. 81-86.

17. MILLER, W., Performing armchair roundoff analyses of statistical algorithms. COMMUNICATIONS IN STATISTICS; SIMULATION AND COMPUTATION, volume B7, no. 3, 1978, pp. 243-255.

18. MORRISON, D.F., MULTIVARIATE STATISTICAL METHODS. McGraw-Hill Book Company, 1967.

19. NARULA, S.C. and J.F. WELLINGTON, Prediction, linear regression and the minimum sum of relative errors. TECHNOMETRICS, 19, no. 2, May 1977, pp. 185-190.

20. PRITCHARD, D.J. and D.W. BACON, Accounting for heteroscedasticity in experimental design. TECHNOMETRICS, 19, no. 2, May 1977, pp. 109-115.

21. SCHEFFÉ, H., THE ANALYSIS OF VARIANCE. John Wiley & Sons, Inc., New York, 4th printing, 1964.

22. SCHEFFÉ, H., Practical solutions of the Behrens-Fisher problem. JOURNAL AMERICAN STATISTICAL ASSOCIATION, 65, pp. 1501-1508.

23. SCHMIDT, P., ECONOMETRICS. Marcel Dekker, Inc., New York, 1976.

24. SWAMY, P.A.V.B., J.S. MEHTA and P.N. RAPPOPORT, Two methods of evaluating Hoerl and Kennard's ridge regression. COMMUNICATIONS IN STATISTICS, THEORY AND METHODS, vol. A7, no. 12, 1978, pp. 1133-1155.

25. THOMPSON, M., Selection of variables in multiple regression: Part I.
    A review and evaluation. INTERNATIONAL STATISTICAL REVIEW, $\underline{46}$, no. 1,
    April 1978, pp. 1-19.

26. VAN DEN BURG, A.J., R.T. VAN DER HAM and J.P.C. KLEIJNEN, GENERALIZA-
    TION OF SIMULATION RESULTS: PRACTICALITY OF STATISTICAL METHODS.
    (In two parts). Reports FEW. 67-68, Department of Economics, Katho-
    lieke Hogeschool, Tilburg (Netherlands), Dec. 1977.

27. WEEKS, J.K. and J.S. FRYER, A simulation study of operating policies
    in a hypothetical dual-constrained job shop. MANAGEMENT SCIENCE,
    $\underline{22}$, no. 12, August 1976, pp. 1362-1371.

ACKNOWLEDGEMENT