Regression analysis for simulation practitioners

Kleijnen, J.P.C.

Published in:
The Journal of the Operational Research Society

Publication date:
1981

J. Opl Res. Soc. Vol. 32, pp. 35 to 43
Pergamon Press Ltd 1981. Printed in Great Britain

0160-5682/81/0101-0035$02.00/0

Regression Analysis for Simulation Practitioners

JACK P. C. KLEIJNEN
School of Economics, Social Sciences and Law, Tilburg University, Netherlands

Some simple extensions of elementary regression analysis useful for analyzing simulation experiments are given. In simulation variance estimates (standard errors) are usually available but often do not satisfy the assumption of constant variance which underlies elementary regression analysis. Two approaches are possible: either switch to Generalized (or Weighted) Least Squares or continue to use Ordinary Least Squares. The consequences of both approaches are surveyed. Testing the adequacy of the regression model is discussed in detail. A case-study illustrates the statistical techniques. Alternatives to Least Squares are briefly indicated.

INTRODUCTION

IN MOST practical and academic simulation studies the experimenter obtains an estimate y of the system response of interest (e.g. mean queuing time) plus the standard error s of this estimate. The standard errors s_i ($i = i, \ldots, N$) of the responses for N different system configurations often show large differences, and hence the assumption of constant variance obviously does not hold. For example, in a case-study s_i^2 ranged from 64 to 93,228. It has become more and more accepted to analyse the outputs of a simulation experiment by using techniques like Analysis of Variance (ANOVA) and regression analysis, ANOVA being just a special case of regression analysis.[1] However, in virtually all practical applications constant variance is assumed.

When conducting a simulation experiment, the investigator has in his mind a list of possibly important factors or variables. He starts out with a tentative regression model; this metamodel formalizes the effects of the factors on the simulation model's response. To estimate these effects a number of system variants specified by the factor combinations, is simulated. From the simulation responses $\mathbf{y}$ and the combinations of variables $\mathbf{X}$ the effects $\boldsymbol{\beta}$ are estimated, using either Ordinary of Generalized Least Squares (OLS, GLS). The resulting regression model is validated using a few additional simulation runs. For the validated regression model, the significance of the various effects is tested. The case-study shows that GLS results are more accurate than OLS results. Readers interested in technical details and additional references may write to the author for the original, unabridged version.

LEAST SQUARES AND HETEROGENEOUS VARIANCES

Ordinary Least Squares uses a strictly mathematical (i.e. non-statistical) criterion: minimize the sum of squared deviations. The resulting estimator is

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}' \cdot \mathbf{X})^{-1} \mathbf{X}' \cdot \mathbf{y}. \tag{1}$$

If the standard statistical assumptions of normally and independently distributed (NID) errors e with constant variance σ^2 and zero expectation, i.e.

$$e \sim \text{NID}(0, \sigma^2) \tag{2}$$

are introduced, then the OLS estimator is known to be the best linear unbiased estimator (BLUE), "best" meaning minimum variance. The covariance-matrix of $\hat{\boldsymbol{\beta}}$ is

$$\boldsymbol{\Omega}_{\hat{\beta}} = \sigma^2 \cdot (\mathbf{X}' \cdot \mathbf{X})^{-1} \tag{3}$$

In practice (3) is applied using standard software. Usually the common variance σ^2 in (3) is estimated from the Mean Squared Residuals (MSR):

$$\text{MSR} = \sum_{1}^{N} (y_i - \hat{y}_i)^2/(N - q) \tag{4}$$

where q denotes the number of estimated parameters. The MSR has only $(N - q)$ degrees of freedon (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. 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 the information about the standard errors could be used to give a more precise extimator of σ^2, if a common variance is assumed.

If the variances are not equal, then the OLS algorithm may still be used, but then (3) does not hold anymore. To derive the correct standard errors of the OLS estimators $\hat{\boldsymbol{\beta}}$, consider a vector of stochastic variables, say $\mathbf{Y}_1$, with covariance matrix $\boldsymbol{\Omega}_1$. Next introduce a linear transformation of $\mathbf{Y}_1$:

$$\mathbf{Y}_2 = \mathbf{A} \cdot \mathbf{Y}_1. \tag{5}$$

Then $\mathbf{Y}_2$'s covariance matrix can be proven[2] to be

$$\boldsymbol{\Omega}_2 = \mathbf{A} \cdot \boldsymbol{\Omega}_1 \cdot \mathbf{A}'. \tag{6}$$

Applying this result to (1), defining for convenience

$$\mathbf{W} \equiv (\mathbf{X}' \cdot \mathbf{X})^{-1} \cdot \mathbf{X}' \tag{7}$$

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

$$\boldsymbol{\Omega}_{\hat{\beta}} = \mathbf{W} \cdot \boldsymbol{\Omega} \cdot \mathbf{W}' \tag{8}$$

where $\boldsymbol{\Omega}$ denotes the covariance matrix of e (or equivalently y). An estimator $\hat{\boldsymbol{\Omega}}_{\hat{\beta}}$ can be easily computed by a computer program that reads the values of the independent variables $\mathbf{X}$ and the estimator $\downarrow \hat{\boldsymbol{\Omega}}$. Obviously the OLS estimator remains unbiased.

Summarizing so far, if the variances are not constant then (3) and (4) are replaced by (8) in which $\boldsymbol{\Omega}$ is estimated from the N individual simulation runs, and $\hat{\boldsymbol{\Omega}}$ becomes a diagonal matrix $\hat{\mathbf{D}}$ with elements s_i^2, each s_i^2 having d_i degrees of freedom.

Note that in simulation the observations y can indeed be made strictly independent through the use of different random numbers in each simulation run (no common or antithetic random numbers). Hence $\boldsymbol{\Omega}$ is reduced to the diagonal matrix $\mathbf{D}$. In the simulation of steady-state behaviour, runs might be continued until each run yields the same *estimated* variance. In practice such an approach is not popular.

If the standard assumptions in (2) do *not* hold, then a BLUE results when GLS is applied:

$$\tilde{\boldsymbol{\beta}} = (\mathbf{X}' \cdot \boldsymbol{\Omega}^{-1} \cdot \mathbf{X})^{-1} \cdot \mathbf{X}' \cdot \boldsymbol{\Omega}^{-1} \cdot \mathbf{y}. \tag{9}$$

The covariance matrix of the GLS estimator is

$$\boldsymbol{\Omega}_{\tilde{\beta}} = (\mathbf{X}' \cdot \boldsymbol{\Omega}^{-1} \cdot \mathbf{X})^{-1}. \tag{10}$$

For independent observations $\boldsymbol{\Omega}$ reduces to the diagonal matrix $\mathbf{D}$, and GLS can be simplified to weighted least squares, the weight for observation y_i being inversely proportional to its variance σ_i^2. However, in practice $\boldsymbol{\Omega}$ or $\mathbf{D}$ is unknown and has to be estimated. Two options are available:

(i) Use OLS even when the classical assumptions of (2) are violated and apply (8).

(ii) Estimate $\boldsymbol{\Omega}$ and substitute the estimator $\hat{\boldsymbol{\Omega}}$ into (9). As Schmidt[3] shows, the resulting estimator has the same asymptotic distribution as the regular GLS estimator and remains unbiased (under mild technical conditions). Unfortunately, its exact small sample behaviour is unknown. Elsewhere[4] a small Monte Carlo experiment is presented includ-

ing the following sampling results:

—GLS with estimated covariance matrix $\hat{\boldsymbol{\Omega}}$ gives point estimators with smaller variances than OLS estimators. This result is intuitively acceptable because OLS yields BLUE only if the variances σ_i^2 are constant; the "estimated GLS" incorporates the information s_i^2 on the actual variances σ_i^2.

—For the "estimated GLS" estimators the standard errors might still be computed through (10), a formula—strictly speaking—valid for known $\boldsymbol{\Omega}$ or for "large" samples. (Intuitively, replacing $\boldsymbol{\Omega}$ by its estimator $\hat{\boldsymbol{\Omega}}$ increases the variance compared to (10)!)

The significance of an estimated regression parameter $\hat{\beta}_j (j = 1, \ldots, q)$ can be tested by the Student t-test:

$$t_d = \frac{\hat{\beta}_j - \beta_j^0}{\sqrt{\hat{\text{var}}\,(\hat{\beta}_j)}}. \tag{11}$$

Here β_j^0 is the hypothesized value, usually zero. The denominator follows from the main diagonal of $\hat{\boldsymbol{\Omega}}_{\hat{\beta}}$. The index d denotes 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. If the postulated value β_j^0 is accepted, the regression model's remaining parameters $\beta_{j'}$ $(j' \neq j)$ can be reestimated.

VALIDATION OF THE 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 $(k \geqslant 1)$. The experimental design fixes x_{i1} through x_{ik} with $i = 1, \ldots, N$. 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.[1] Deciding which interactions to include in $\mathbf{X}$ specifies the form of the metamodel which is linear in its parameters $\boldsymbol{\beta}$:

$$\mathbf{y} = \mathbf{X} \cdot \boldsymbol{\beta} + \mathbf{e}. \tag{12}$$

If (12) is a good approximation, using estimators for its parameters $\boldsymbol{\beta}$ yields an accurate predictor $\hat{\mathbf{y}}$. This predictor can be checked against the outcome of an actual simulation run y. More precisely, let $\mathbf{x}_{N+1}$ denote the column vector of prespecified values of the independent variables in a new simulation run, i.e. this run was not used in computing the estimator $\hat{\boldsymbol{\beta}}$, in other words $\mathbf{x}_{N+1}$ is not included in $\mathbf{X}$. Hence the expected value of the simulation output is predicted by

$$\hat{y}_{N+1} = \mathbf{x}'_{N+1} \cdot \hat{\boldsymbol{\beta}}. \tag{13}$$

Using (6) yields

$$\text{var}\,(\hat{y}_{N+1}) = \mathbf{x}'_{N+1} \cdot \boldsymbol{\Omega}_{\hat{\beta}} \cdot \mathbf{x}_{N+1} \tag{14}$$

where $\boldsymbol{\Omega}_{\hat{\beta}}$ is given in (8) or (10). The simulation program reads $\mathbf{x}_{N+1}$ and yields the output y_{N+1} with its estimated variance s_{N+1}^2, based on d_{N+1} degrees of freedom. The model's validity can be tested through a Student t-statistic:

$$t_d = \frac{\hat{y}_{N+1} - y_{N+1}}{[\hat{\text{var}}\,(\hat{y}_{N+1}) + \hat{\text{var}}\,(y_{N+1})]^{1/2}} \tag{15}$$

where d (the d.f. of t) may be set to the minimum of the d.f. of $\hat{\text{var}}\,(\hat{y}_{N+1})$ and $\hat{\text{var}}\,(y_{N+1})$, resulting in a conservative test, i.e. the actual type I error may be smaller than the nominal α-value.[5]

If the constant-variance assumption holds, then an F-test for lack-of-fit is possible. This test compares the estimators s_i^2 to the Mean Squared Residuals of (4). Apart from its restrictive assumptions, its power (inverse of β-error) is low, when its d.f. are small. Note that authors disagree about the sensitivity of the F-test to heterogeneity of variance and to nonnormality.

After a validation run is accepted, it can be added to $\mathbf{X}$ and $\mathbf{y}$ so that $\boldsymbol{\beta}$ can be estimated more precisely. It seems wise to have $\mathbf{x}_{N+1}$ correspond with the centre of the design (i.e. the quantitative factors satisfy $x = 0$) to test quadratic effects. Some validation runs ($N + 1, N + 2, \ldots$) should correspond to x-values occuring in practice, because the use of experimental designs to specify $\mathbf{X}$ means that the x-values correspond to reasonable extreme conditions rather than to common conditions. A trick to obtain validation runs is to delete one run i from the N old observations, yielding $\mathbf{y}^{(i)}$ and $\mathbf{X}^{(i)}$ and to use $\mathbf{y}^{(i)}$ and $\mathbf{X}^{(i)}$ to compute $\hat{\boldsymbol{\beta}}^{(i)}$. The $\hat{\boldsymbol{\beta}}^{(i)}$ can be used to predict y_i.

SIMULTANEOUS TESTS

Regression analysis involves a number of tests, for 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 in the next section, ten extra runs are available to test the adequacy of the regression (meta)model. By definition the α-error implies

$$P(t_d \geqslant t_d^a | H_0) = \alpha. \tag{16}$$

Hence even if the null-hypothesis of an adequate model holds, 10 validation 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 (16) by α/n where n is the number of tests, i.e. $n = 10$. Instead of this simple Bonferroni approach more complicated "multiple comparison procedures" are available. Note that protection of the α-error increases the β-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:

$$y_i = \beta_0 + \sum_{j=1}^{k} \beta_j x_{ij} + e_i. \tag{17}$$

Then the parameters β_j can be tested through the t-test of (11). Each factor is considered individually, i.e. the interpretation of the experiment does not hinge on the joint results of the tests. Therefore the familiar "per comparison" error rate of, say, $\alpha = 10\%$ is recommended. Remember that in the validation phase the model is rejected if any validation run yields a significant t-value, i.e., the experimentwise error rate is then relevant.

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

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_{12} x_{i1} x_{i2} + \beta_{11} x_{i1}^2 + \beta_{22} x_{i2}^2 + e_i. \tag{18}$$

Suppose that the t-test of (11) shows that all $\hat{\beta}$'s are significant except for $\hat{\beta}_{11}$. Remember, that $\hat{\beta}_{11}$ is an unbiased estimator of β_{11}; if the assumptions of (2) hold, then $\hat{\beta}_{11}$ is a BLUE. Strong reasons may exist to formulate a null-hypothesis. For instance, in (17) the parsimonious character of scientific models requires that instead of postulating that "everything depends on everything else", the observation y should be explained by as few factors as possible: $H_0^{(j)}: \beta_j = 0 \, (j = 1, \ldots, k)$. Equation (18), however, postulates that y is a quadratic polynomial in x_1 and x_2. Hence a small, but non-zero, value of $\hat{\beta}_{11}$ should be maintained rather than set to zero.

A different question may arise: can (18) 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:[6]

(1) Formulate the *composite* hypothesis—

$$H_0: \beta_{12} = 0 \wedge \beta_{11} = 0 \wedge \beta_{22} = 0 \tag{19}$$

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, i.e. 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.

(2) The hypothesis of (19) can also be tested by applying the individual t-tests of (11) with α replaced by $\alpha/3$: Bonferroni approach.

(3) A cruder approach estimates the first-order polynomial

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + e_i \tag{20}$$

and validates this model with runs not used in estimating (20); see (15). This alternative is cruder, because if the simpler model of (20) is rejected, it is unknown whether this rejection is caused by a large value for β_{12}, for β_{11} or for β_{22}.

AN APPLICATION

This section summarizes a case study presented in detail elsewhere[7] (the previous publication includes some erroneous Monte Carlo results.) Europe Container Terminus (ECT) in the Rotterdam harbour provides facilities for handling and storing containers. A simulation model represents storage capacity w as a function of yearly throughput (production). A given amount of annual production 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_6. Every 8 simulated hours, the simulation gives a snapshot of the storage size. From this time series w_t ($t = 1, \ldots, T$) a frequency diagram is formed. The frequency diagram yields an average and a few selected quantiles such as the 90% quantile. Figure 1 is a simplified flowchart 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 are analyzed similarly, although more sophisticated multivariate analysis would be better.

The complicated simulation model of Figure 1 defines a function f:

$$y = f(x_1, \ldots, x_6, \mathbf{r}) \tag{21}$$

where $\mathbf{r}$ is the random number vector. The complicated function f is approximated (in the area of experimentation) by a regression model linear in its parameters β but not necessarily linear in the variables x. Preliminary studies suggested that the response y reacts nonlinearly to the interarrival time but linearly to the interarrival rate; therefore a simple transformation $1/x$ simplifies the 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 $\boldsymbol{\beta}$ comprises one overall mean β_0, six main effects β_1 through β_6, and six interactions, altogether $q = 13$ parameters. The selection of an appropriate $\mathbf{X}$ is in the domain of experimental design theory.[1,6] Application of this theory results in a 16 by 13 X-matrix. (Readers familiar with experimental design techniques can construct $\mathbf{X}$ by using the generators $\mathbf{1} = \mathbf{56}$ and $\mathbf{3} = \mathbf{45}$.) So 3 degrees of freedom remain for a possible F-test for lack-of-fit. However, instead of this F-test the t-test of (15) is applied to ten extra runs executed in addition to the above 16 runs.

In Table 1 the standard errors for the GLS estimates $\hat{\beta}$ and hence the corresponding t-values, are based on the asymptotic formula (10). In Table 2 vâr (y) can be computed after dividing each simulation run into nine subruns. Table 2 shows that the OLS regression model need not be rejected, since the maximum absolute value of the 10 t-statistics is 1.67 whereas the significance level is 2.33 for $\alpha = 0.20/10$ (experimentwise error of 20%). For GLS the validation runs need not be rejected either (not shown in tables).

After accepting the regression model, the 10 validation runs are included in $\mathbf{X}$ and $\mathbf{y}$, and $\boldsymbol{\beta}$ 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 to $t = 64$ and 7.9.

Note that the experimental design matrix, say $\mathbf{Z}$, consists of standardized variables ($z = +1$ or $z = -1$), whereas the actual design and the regression model contains "user" variables x, e.g. x_1 is either 200 or 1000. The user variables have as significant regression

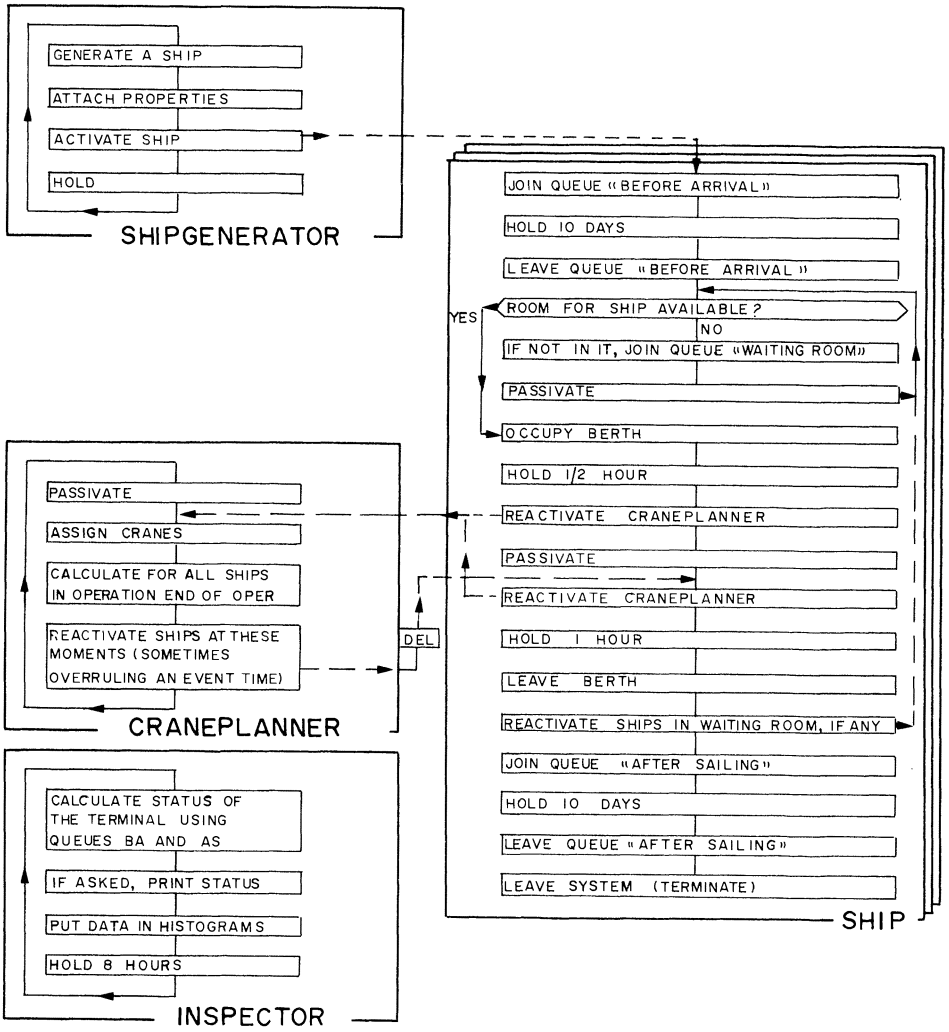

FIG. 1. *The simulation model.*

TABLE 1. OLS AND GLS ESTIMATOR OF REGRESSION PARAMETERS β BASED ON $2^{8-4} = 16$ RUNS

Indexes on β	$\hat{\beta}$	s_{β}	t	$\tilde{\beta}$	s	t
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

* Significant at $\alpha = 0.025$.
† Significant at any $\alpha > 0.00005$.

TABLE 2. MODEL VALIDATION (OLS)

y	$\hat{y}$	$y - \hat{y}$	vâr (y)	vâr $(\hat{y})$	t
8332	8715	−383	22.102	30,494	−1.67
3002	2919	83	1156	5092	1.05
729	743	−14	544	964	−0.36
1725	1774	−49	625	1142	−1.16
1893	1814	79	4444	1205	1.05
685	684	1	107	847	0.03
2977	3058	−81	4761	8308	−0.71
8469	8415	54	10,885	20,808	0.30
608	595	13	152	920	0.40
1674	1624	50	514	1138	1.23

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, where γ denotes the parameters of the standardized variables z).

Summarizing, some parameter estimates $\hat{\beta}$ were found to be insignificant, after validating the first 16 runs using ten extra runs, and then reestimating $\boldsymbol{\beta}$ from all 26 runs. Next these insignificant parameters are set to zero, and the remaining β's, i.e. β_{23} and β_{12}, are again reestimated.

In general, one should examine the residuals $y - \hat{y}$ to see whether they satisfy the classical assumptions of (2).[8] Studying the responses (especially the residuals) and apply-

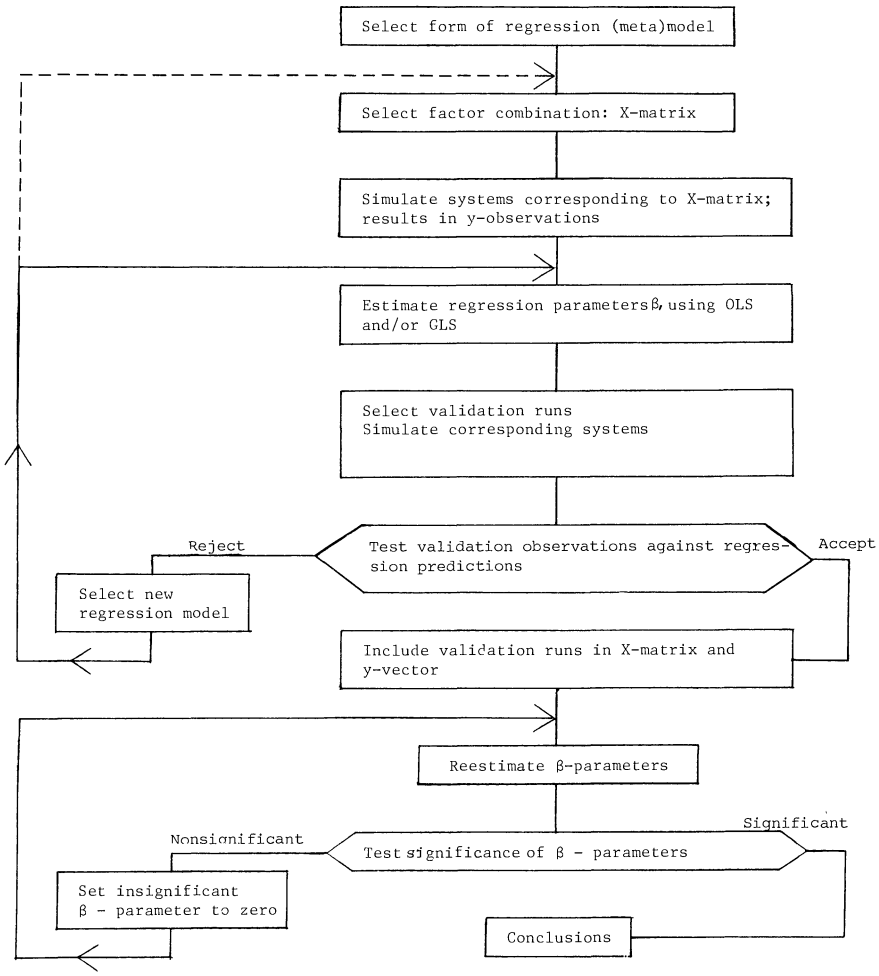

FIG. 2. *Summary of regression procedure.*

ing common sense to them, revealed certain patterns that suggest the importance of interactions until then ignored, namely β_{14} and β_{15}. Fortunately, incorporating these two new effects into **X** left **X** non-singular (see also next section). The resulting $\hat{\boldsymbol{\beta}}$ still contains as significant parameters only $\hat{\beta}_{23}$ and $\hat{\beta}_{12}$.

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

The above procedure is summarized in Figure 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.[9] However, specifying the regression model involves more than a bag of statistical tricks; it also requires intuition and prior knowledge based on relevant theories and empirical data. In the present case study the most significant parameter β_{12} was the one parameter suggested by a simplified analytical model.

ALTERNATIVES TO OLS AND GLS

Both OLS and GLS use as their criterion the minimization of squared residuals. Simulation practitioners tend to focus on *relative* absolute residuals $|y - \hat{y}|/y$. The absolute errors $|y - \hat{y}|$ lead to a linear programming problem.[10] Unfortunately, the properties of the latter 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 of y or x. Robust regression estimators are surveyed in the references.[11,12]

If the X-matrix is *ill-conditioned*, ridge estimation may be of interest, i.e. the estimators of $\boldsymbol{\beta}$ are no longer unbiased; however, their bias may be outweighted by a decrease in variance attained through a proper choice of the ridge algorithm parameters.[9] In simulation **X** might 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.

Dempster *et al.*[13] 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 to outliers is minimized.[14]

One more alternative is provided by the Bayesian decision-theoretic model: prior probabilities on parameters like $\boldsymbol{\beta}$ are postulated (Bayes approach), together with loss functions like $\Sigma w_i (\beta_j - \hat{\beta}_j)^2$. Instead of fixing the α-errors, the expected *a posteriori* (after taking the sample) loss is minimized, or the maximum loss is minimized.[13]

CONCLUSION

To mitigate the *ad hoc* character of simulation, regression analysis can be used to produce a metamodel. The metamodel aids in interpreting the simulation results.

The regression analysis can use OLS or GLS. When applying OLS the experimenter should check for nonconstant variances σ_i^2 (estimated from the individual simulation runs). When variances change from run to run, the formula for $\boldsymbol{\Omega}_{\hat{\beta}}$ (the covariance matrix of the estimated parameters $\hat{\beta}$) is affected which changes the corresponding t-test for significance. A Monte Carlo experiment suggested that GLS with estimated $\boldsymbol{\Omega}$ (covariance matrix of the observations) results in a covariance matrix for the β-estimators that can be approximated accurately by the asymptotic formula (10).

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

The form of the model and the values specified in nullhypotheses have to come from nonstatistical 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.

REFERENCES

[1] J. P. C. KLEIJNEN (1979) The role of statistical methodology in simulation. In *Methodology in Systems Modelling and Simulation* (B. P. ZEIGLER *et al.*, Ed.). North-Holland, Amsterdam.
[2] H. SCHEFFÉ (1964) *The Analysis of Variance*. Wiley, New York.
[3] P. SCHMIDT (1976) *Econometrics*. Marcel Dekker, New York.
[4] J. P. C. KLEIJNEN, R. BRENT and R. BROUWER (1980) Small-sample behavior of weighted least squares in experimental design applications. *Comm. Stud. Sim. Comp.* (forthcoming).
[5] H. SCHEFFÉ (1970) Practical solutions of the Behrens-Fisher problem. *J. Am. statist. Ass.* **65,** 1501–1508.
[6] J. P. C. KLEIJNEN (1975) *Statistical Techniques in Simulation*. Marcel Dekker, New York.
[7] J. P. C. KLEIJNEN, A. J. VAN DEN BURG and R. T. VAN DER HAM (1979) Generalization of simulation results: practicality of statistical methods. *Eur. J. Opl Res.* **3,** 50–64.
[8] S. R. WILSON (1979) Examination of regression residuals. *Aust. J. Statist.* **21,** 18–29.
[9] R. R. HOCKING (1976) The analysis and selection of variables in linear regression. *Biometrics* **32,** 1–49.
[10] J. E. GENTLE (Ed.) (1977) Special issue on computations for least absolute values estimation. *Comm. Stat.* **B6,** 313–446.
[11] L. DENBY and W. A. LARSEN (1977) Robust regression estimators compared via Monte Carlo. *Comm. Stat.* **A6,** 335–362.
[12] R. W. HOGG (1977) Robustness, Special Issue of *Comm. Stat.* **A6,** 789–894.
[13] A. P. DEMPSTER, M. SCHATZOFF and N. WERTMUTH (1977) A simulation study of alternatives to ordinary least squares. *J. Am. statist. Ass.* **72,** 77–106.
[14] G. E. P. BOX and N. R. DRAPER (1975) Robust designs. *Biometrika* **62,** 347–352.

References

[9] **A Biometrics Invited Paper. The Analysis and Selection of Variables in Linear Regression**
R. R. Hocking
Biometrics, Vol. 32, No. 1. (Mar., 1976), pp. 1-49.
Stable URL:
http://links.jstor.org/sici?sici=0006-341X%28197603%2932%3A1%3C1%3AABIPTA%3E2.0.CO%3B2-P

[14] **Robust Designs**
George E. P. Box; Norman R. Draper
Biometrika, Vol. 62, No. 2. (Aug., 1975), pp. 347-352.
Stable URL:
http://links.jstor.org/sici?sici=0006-3444%28197508%2962%3A2%3C347%3ARD%3E2.0.CO%3B2-K