

ornl

RECEIVED

FEB 21 1996

OSTI

**OAK RIDGE
NATIONAL
LABORATORY**

**Heteroscedasticity Generated by Errors
in Predictors**

MARTIN MARIETTA

Darryl J. Downing
Valerii V. Fedorov
Randy K. Nanstad

MANAGED BY
MARTIN MARIETTA ENERGY SYSTEMS, INC.
FOR THE UNITED STATES
DEPARTMENT OF ENERGY

DISTRIBUTION OF THIS DOCUMENT IS UNLIMITED

This report has been reproduced directly from the best available copy.

Available to DOE and DOE contractors from the Office of Scientific and Technical Information, P.O. Box 62, Oak Ridge, TN 37831; prices available from (615) 576-8401, FTS 626-8401.

Available to the public from the National Technical Information Service, U.S. Department of Commerce, 5285 Port Royal Rd., Springfield, VA 22161.

This report was prepared as an account of work sponsored by an agency of the United States Government. Neither the United States Government nor any agency thereof, nor any of their employees, makes any warranty, express or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise, does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government or any agency thereof. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government or any agency thereof.

Computer Science and Mathematics Division
Mathematical Sciences Section

**HETEROSCEDASTICITY GENERATED BY ERRORS IN
PREDICTORS**

Darryl J. Downing ^a
Valerii V. Fedorov ^a
Randy K. Nanstad ^b

^a Mathematical Sciences Section
Computer Science and Mathematics Division
Oak Ridge National Laboratory
P. O. Box 2008
Oak Ridge, Tennessee 37831-6367

^b Metals and Ceramics Division
Oak Ridge National Laboratory
P. O. Box 2008
Oak Ridge, Tennessee 37831-6151

Date Published: November 1995

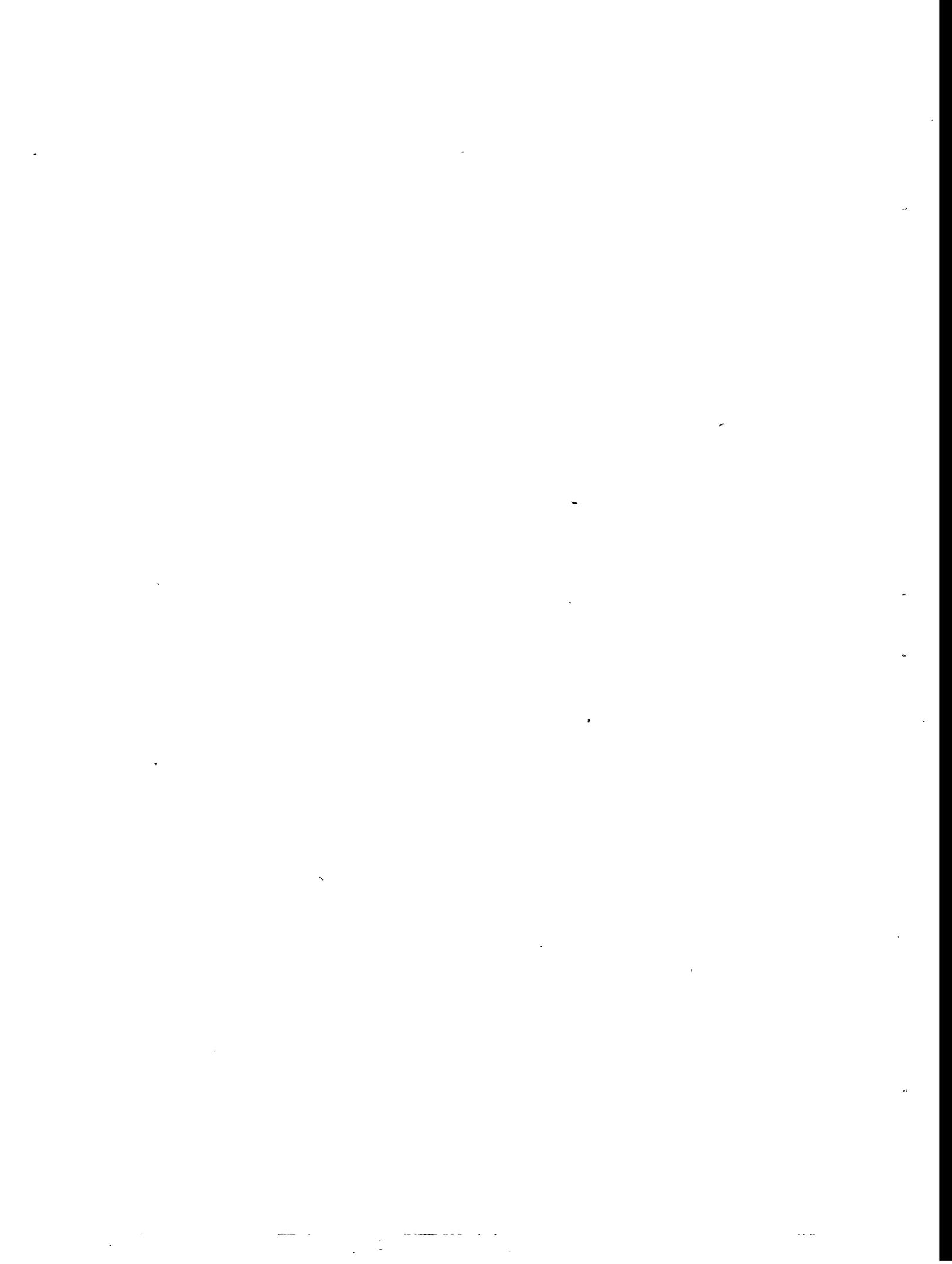
Research sponsored by the Laboratory Directed Research and Development
Program of Oak Ridge National Laboratory

Prepared by the
Oak Ridge National Laboratory
Oak Ridge, Tennessee 37831
managed by
Lockheed Martin, Inc.
for the
U.S. DEPARTMENT OF ENERGY
under Contract No. DE-AC05-84OR21400

MASTER

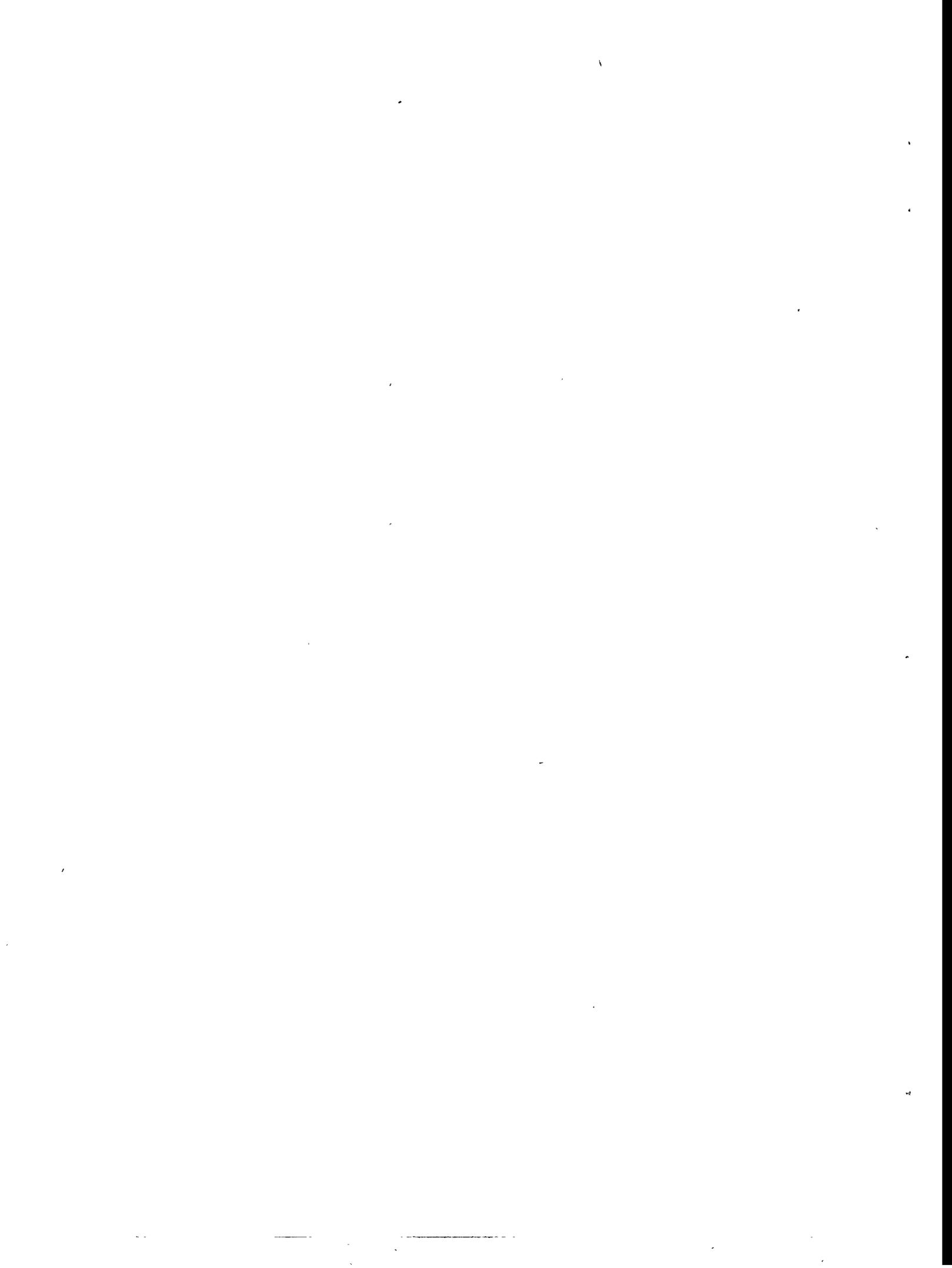
DISTRIBUTION OF THIS DOCUMENT IS UNLIMITED

at



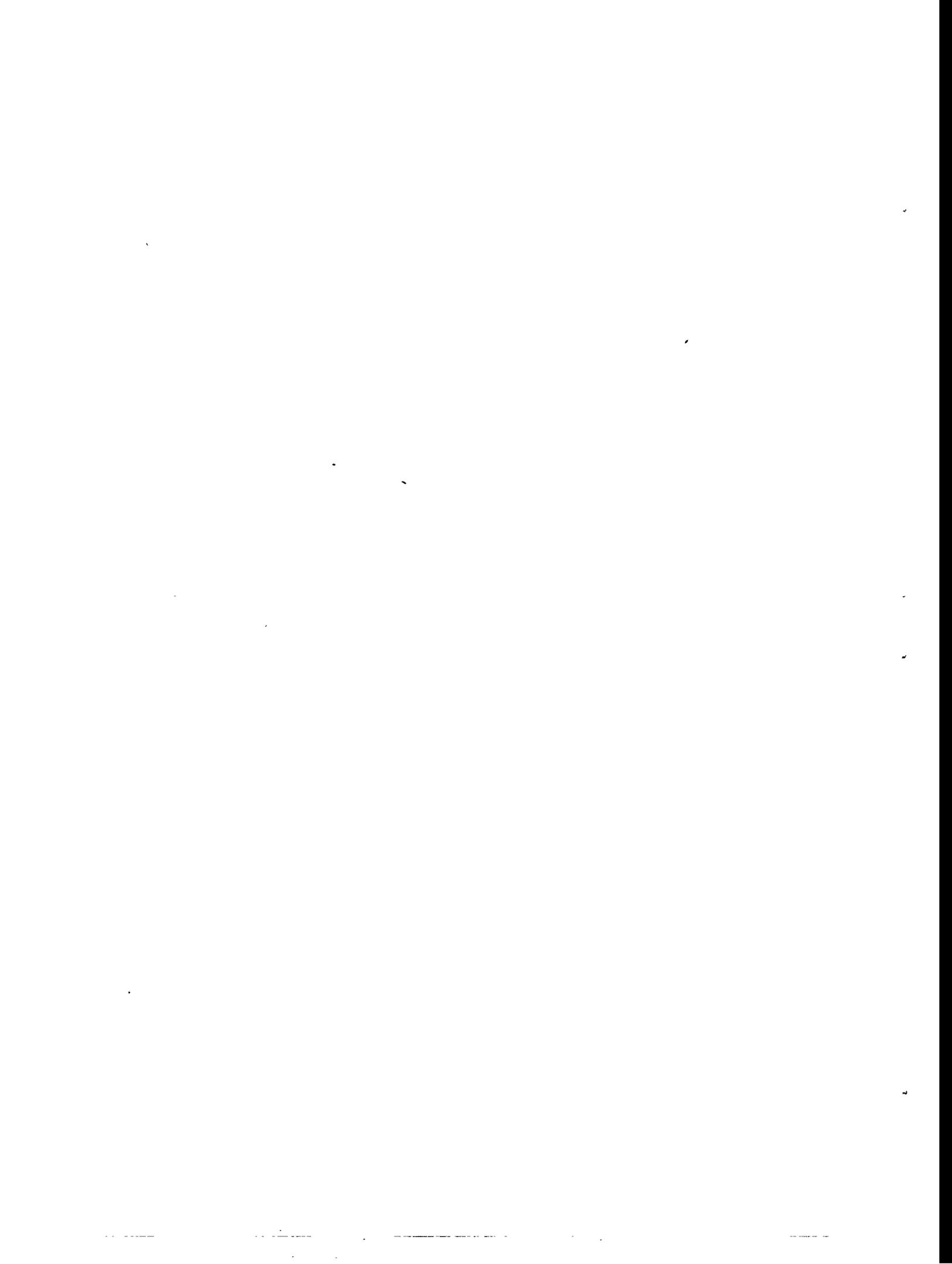
Contents

1	Introduction	1
2	Original regression problem	1
3	Errors in predictors	3
4	Estimation method	9
5	Unknown σ and δ	18
6	Conclusions	27
7	References	29



List of Figures

1	Response function for the 72W irradiated data from DHN 1990. . .	4
2	Response function for the 72W unirradiated data from DHN 1990.	5
3	Response function for the 73W irradiated data from DHN 1990. . .	6
4	Response function for the 73W unirradiated data from DHN 1990.	7
5	Response function for the 72W irradiated data. Calculated with and without adjusting for error in temperature.	13
6	Response function for the 72W unirradiated data. Calculated with and without adjusting for error in temperature.	14
7	Response function for the 73W irradiated data. Calculated with and without adjusting for error in temperature.	15
8	Response function for the 73W unirradiated data. Calculated with and without adjusting for error in temperature.	16
9	Standardized residuals for the 72W irradiated data. Not adjusted for error in temperature.	19
10	Standardized residuals for the 72W irradiated data. Adjusting for error in temperature.	20
11	Standardized residuals for the 72W unirradiated data. Not ad- justed for error in temperature.	21
12	Standardized residuals for the 72W unirradiated data. Adjusting for error in temperature.	22
13	Standardized residuals for the 73W irradiated data. Not adjusted for error in temperature.	23
14	Standardized residuals for the 73W irradiated data. Adjusting for error in temperature.	24
15	Standardized residuals for 73W unirradiated data. Not adjusted for error in temperature.	25
16	Standardized residuals for the 73W unirradiated data. Adjusting for error in temperature.	26



HETEROSCEDASTICITY GENERATED BY ERRORS IN PREDICTORS

Darryl J. Downing

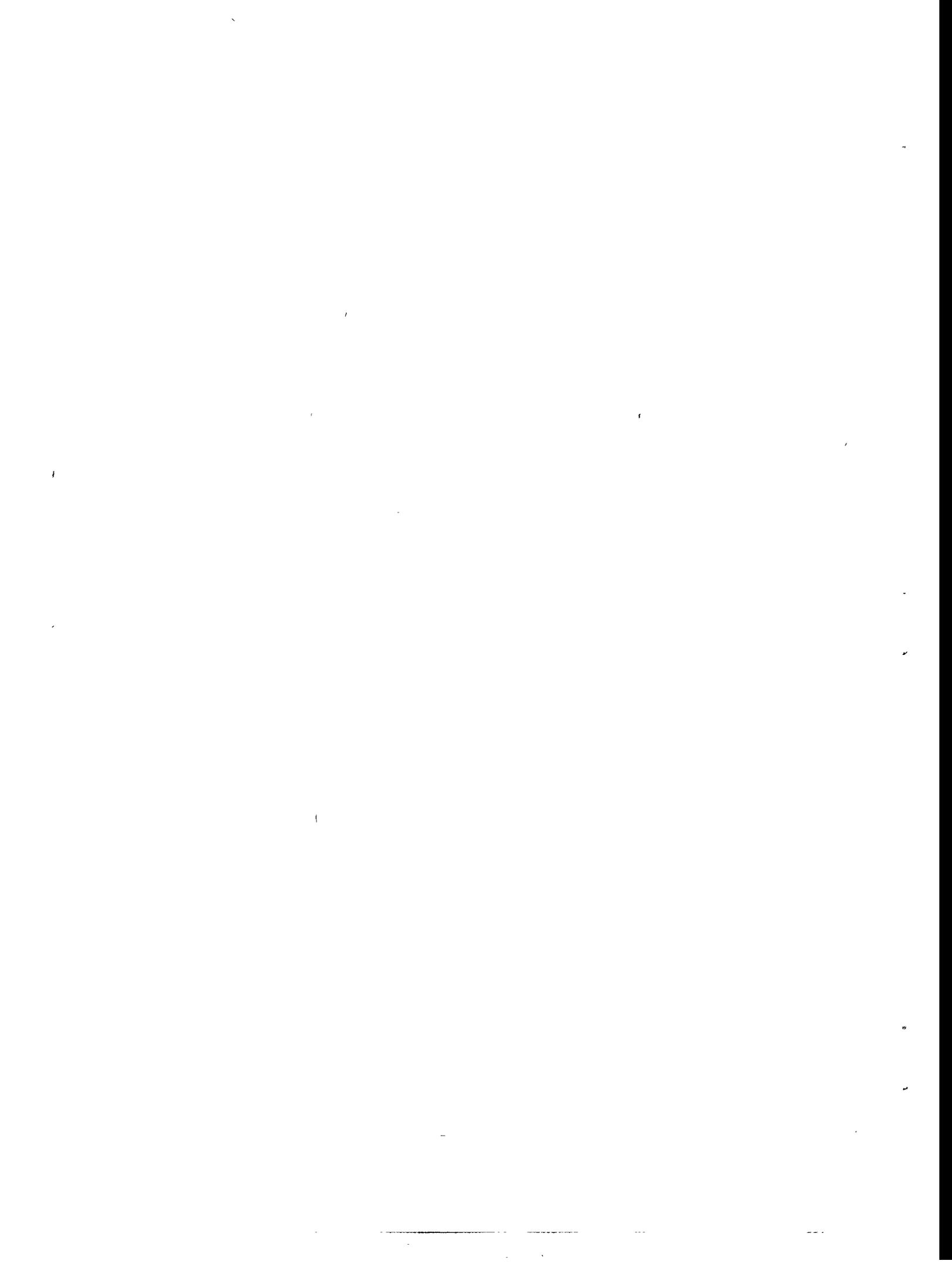
Valerii V. Fedorov

Randy K. Nanstad

Abstract

The heteroscedasticity or changing variance observed in "raw" data may be the result of randomness or uncertainty in the predictor variables. As an example we consider "Charpy Test" experiments widely used to characterize the ductile-brittle toughness of steels such as those used for nuclear pressure vessels. While this type of experiment is of interest in itself, our main objective is to show that the use of a proper statistical technique may help to avoid the use of more complicated physical models to explain the heteroscedasticity of the observations. We also extend the existing method of regression analysis with errors in controllable variables to the case when the variances of the response and the controllable variables are both unknown.

Key words: Predictors subject to errors, regression analysis, errors in controllable variables, iterated estimators, logistic response model.



1. Introduction

This short study was essentially motivated by results reported by Downing, Haggag, and Nanstad , 1990 or in the sequel DHN, 1990, for brevity. They were interested in the analysis of experiments related to the fracture toughness of commercial nuclear light water reactor pressure vessels. One of the tests conducted for toughness determination is the "Charpy" impact test. The tests were conducted over a range of temperatures and the impact energy required to break either irradiated or unirradiated specimens was measured and then modeled as a function of temperature.

In Section 2 we briefly discuss the model used by DHN, 1990 and the results of their statistical analyses. The basic statistical model that includes possible errors in the controllable variables is formulated and discussed in Section 3. Section 4 is based on the approach developed by one of the authors. In Section 5 a more general technique is developed and then applied to the case-study example.

2. Original regression problem

In what follows we accept the basic physical ideas used by DHN, 1990. For obvious reasons the same data sets will be used to make the results comparable.

DHN, 1990, fit the following model

$$y_i = \eta(t_i, \theta) + \sigma \epsilon_i, \quad i = 1, \dots, k, \quad (1)$$

where

y_i is the natural logarithm of energy absorbed at test,

$\eta(t_i, \theta)$ is a given function,

θ are unknown parameters,

ϵ_i are standardized random variables (errors of observations) with the density function

$$p(\epsilon_i) = \exp(\epsilon_i - \exp \epsilon_i), \quad (2)$$

which is widely used in reliability analysis to describe the distribution of extrema. The parameter σ is usually unknown. In DHN, 1990, the response function $\eta(t_i, \theta)$ was approximated by a fourth degree polynomial. Using a fourth degree polynomial in temperature over the range from -100 to 100 C may lead to unstable estimation. Therefore in this study our analysis is based on the logistic response function:

$$\eta(t_i, \theta) = \ln \left[\theta_1 + \frac{\theta_2}{1 + \exp[\theta_3(t - \theta_4)]} \right]. \quad (3)$$

This function is flexible enough and has a more "physical" interpretation than the polynomial model. Data analysis in DHN, 1990, was based on the maximum likelihood method using the error distribution given in (2). For small and moderate σ the maximum likelihood estimators for problem (1) - (3) may be well approximated by the least squares estimators for the following regression model:

$$y_i = \eta(t_i, \theta) + \sigma\mu + \sigma\nu_i \quad (4)$$

where $\nu_i = \epsilon_i - \mu$, $\mu = E(\epsilon_i) = 0.577$ and $\eta(t, \theta)$ is defined by (3).

Moreover, asymptotically ($n \rightarrow \infty$) both estimators coincide. The correction term $\sigma\mu$ may be incorporated in θ_1 and θ_2 through the obvious transformation. Therefore, for the sake of simplicity, we will use the model

$$y_i = \eta(t_i, \theta') + \sigma\nu_i. \quad (5)$$

In what follows the superscript "r" will be omitted when it is not misleading.

Model (1) - (3) and subsequently (5) assumes homogeneous observational errors ($\sigma \equiv \text{constant}$). Numerous physical models discussed in applied publications

(see, for instance, Helm, 1995) support this assumption. The results reported in DHN, 1990, were based on this assumption.

However, one can notice from plots presented in DHN, 1990, and partly cited here (see Fig 1) that the residuals are visibly greater around the mid-point of the so-called ductile to brittle (i.e., high energy to low energy) transition temperature region in each case. In the same temperature range the variability of the response variable is also highest.

The materials in question are known to be anisotropic and heterogenous in terms of microstructural features (e.g., grain size and orientation, etc). Furthermore, in the transition temperature region, there is competition between the separate micromechanisms which lead to brittle or ductile fracture. These factors contribute to variations in measured mechanical properties. Besides those metallurgical contributions to variability, there are two other possible explanations for the observed variability of the response variable. First, it can be that the selected model is not very accurate and the “intrinsic” or “physical” variability of the observed response depends upon the temperature of a tested specimen. The second possibility is that some additional variability induced by an experimenter (or experimental system) may be an explanation of the observed phenomenon. In this paper we will try to explore the second case, suspecting that the variability may arise due to error in temperature control of the test.

3. Errors in predictors

Let us consider the following modification of model (5):

$$\begin{aligned} y_i &= \eta(t_i, \theta) + \sigma \nu_i, \\ t_i &= t_{0i} + \delta \gamma_i, \end{aligned} \tag{6}$$

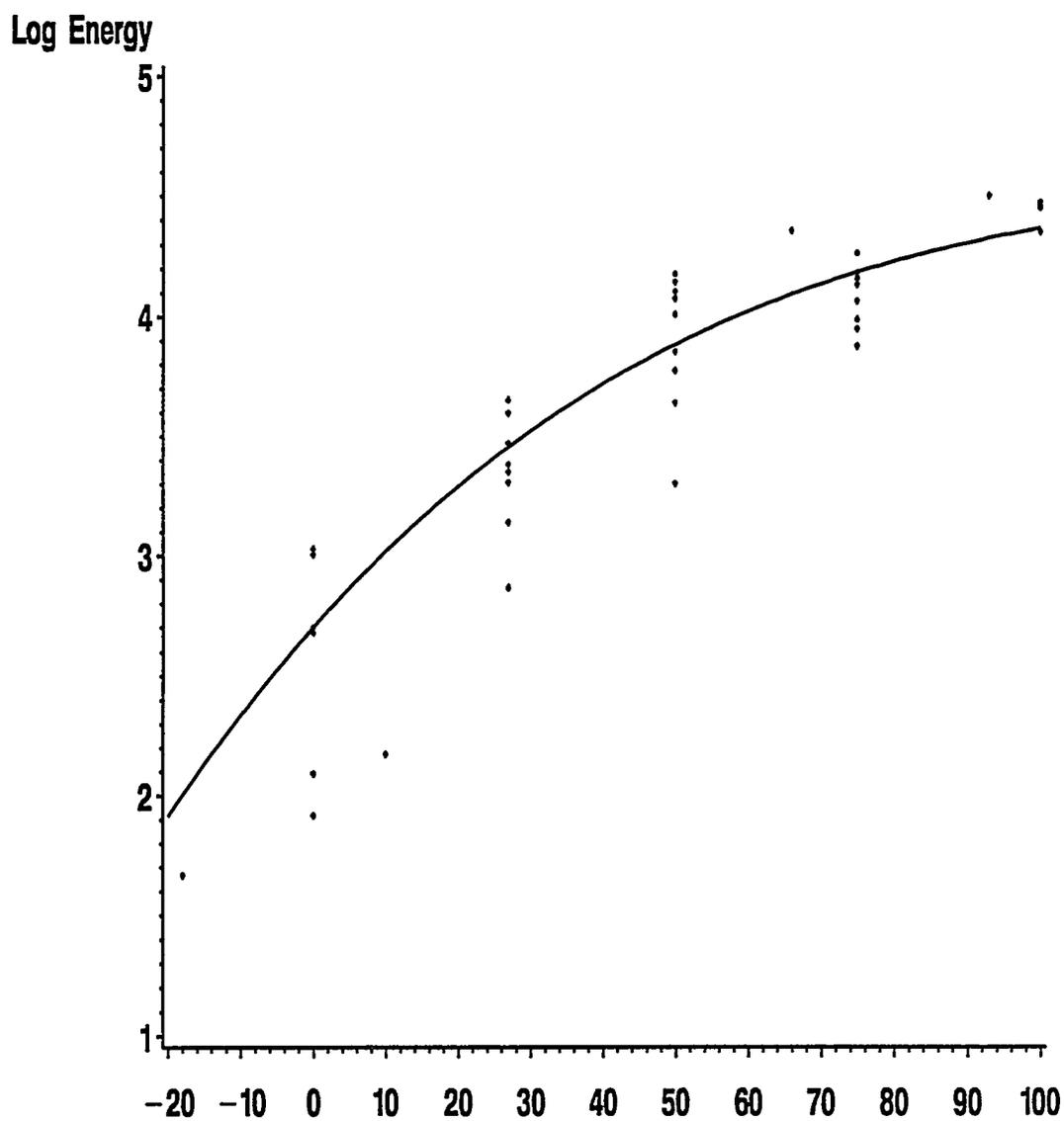


Figure 1: Response function for the 72W irradiated data from DHN 1990.

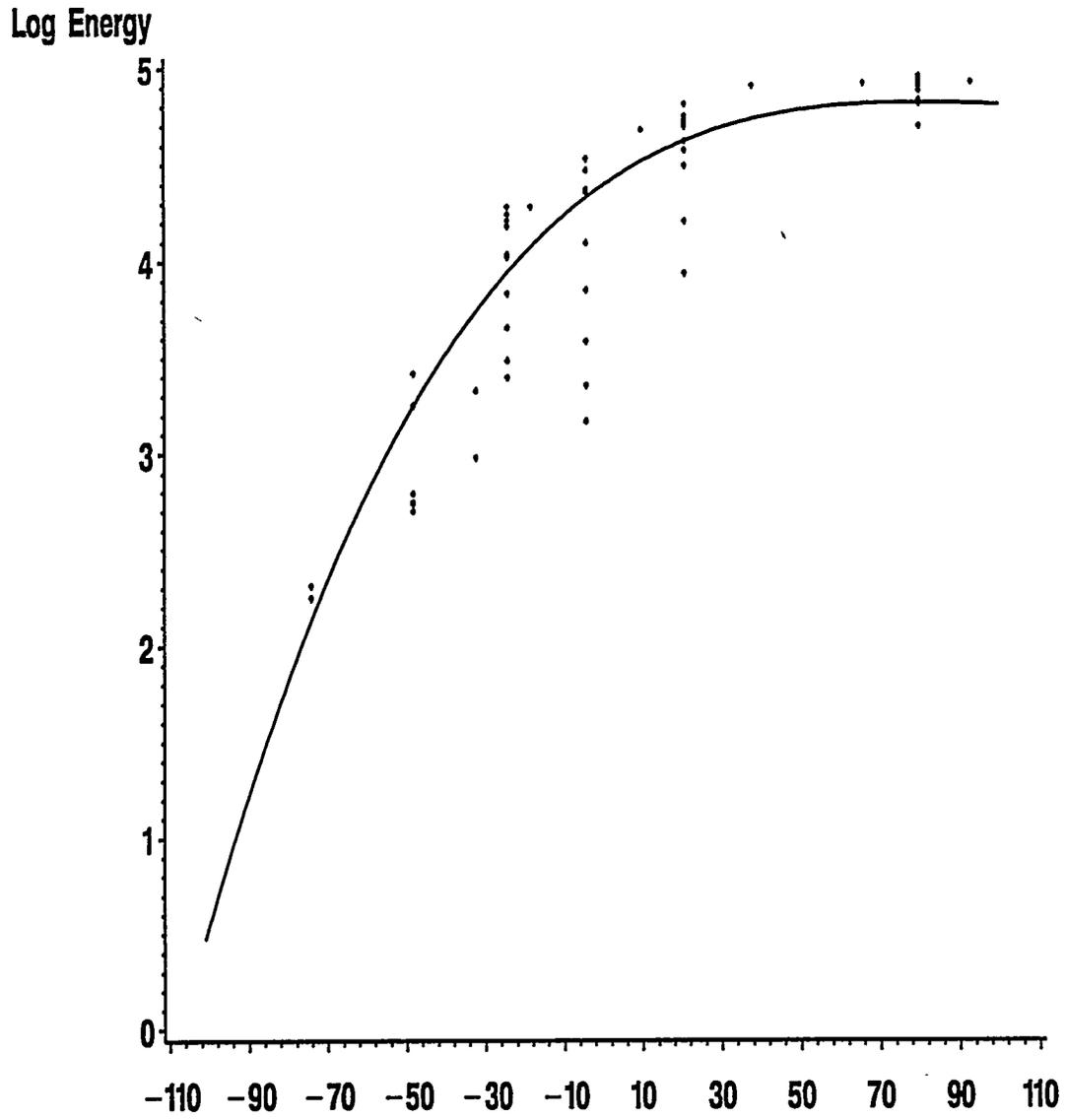


Figure 2: Response function for the 72W unirradiated data from DHN 1990.

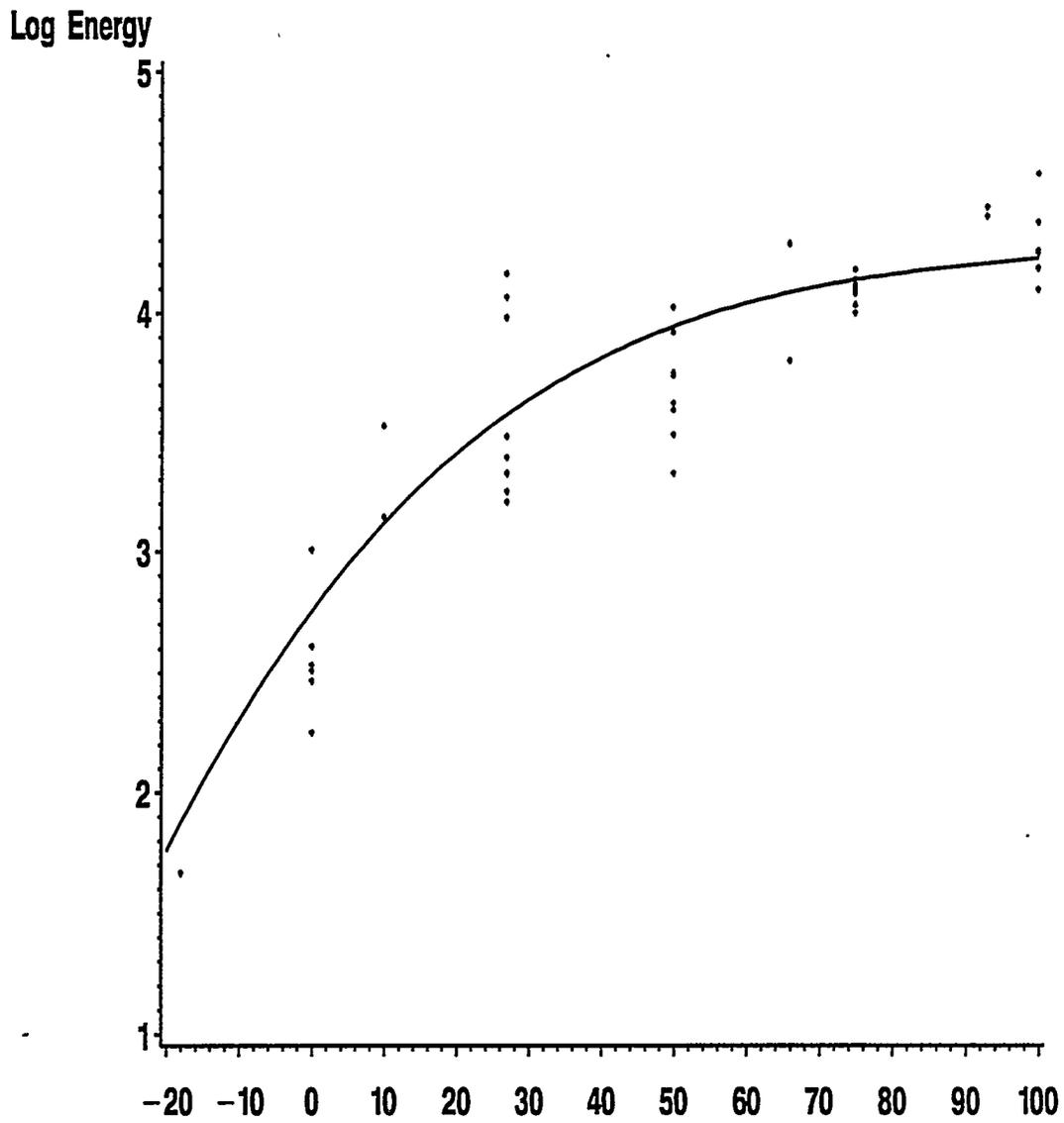


Figure 3: Response function for the 73W irradiated data from DHN 1990.

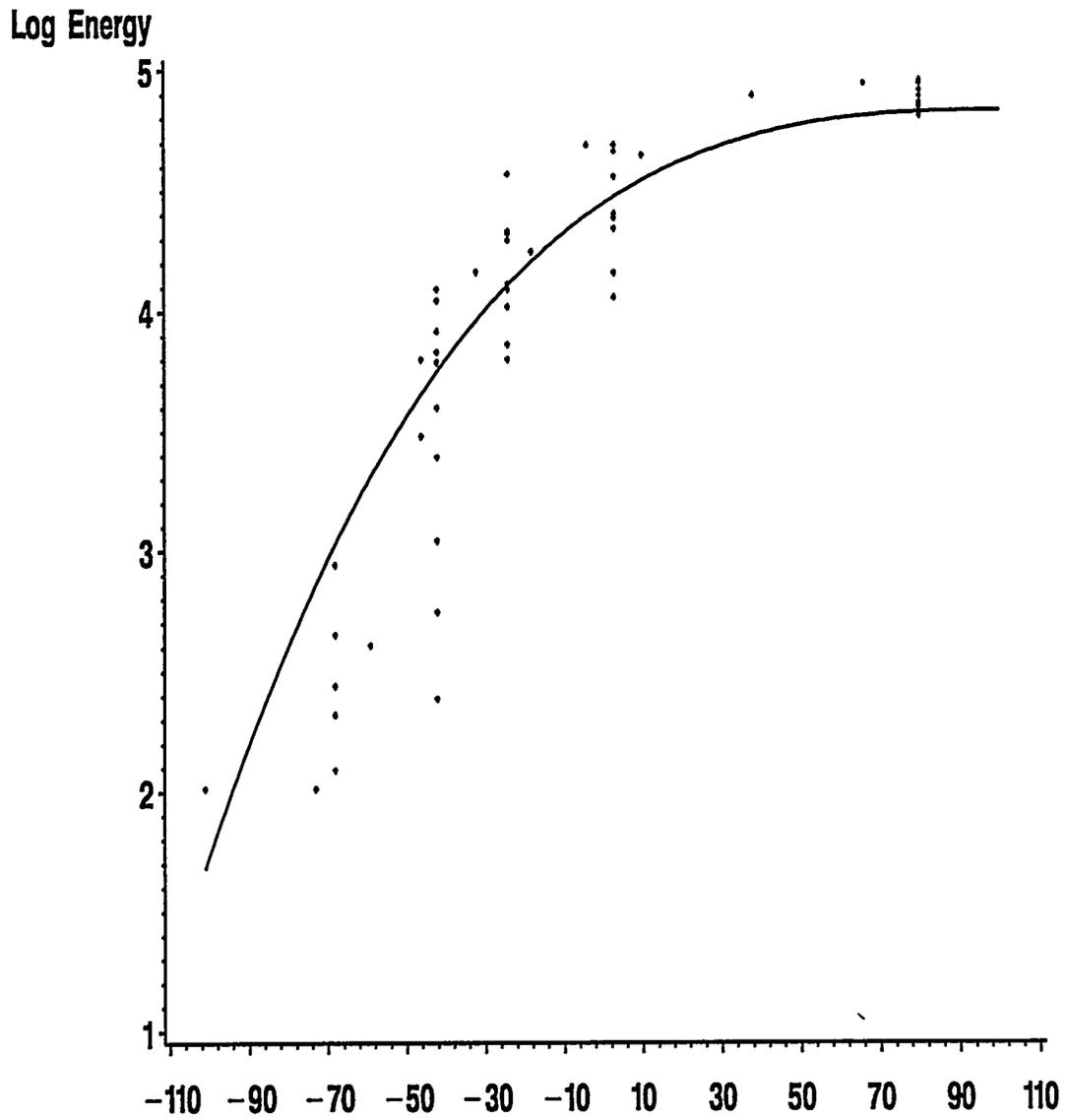


Figure 4: Response function for the ^{73}W unirradiated data from DHN 1990.

where y_i and t_{0i} (the desired temperature) are known, ν_i and γ_i are standardized random variables with zero means and unit variances. We assume that ν_i and γ_i are independent and all random variables corresponding to different observations are independent as well. Model (6) means that an experimenter wants to test specimens at temperatures t_{0i} but cannot control those temperatures completely and therefore the actual tests take place at temperatures $t_{0i} + \delta\gamma_i$. The regression problem for this model was intensively studied for quite a while. If one suspects that the shape of the probability distributions of ν_i and γ_i may significantly influence the estimates then the conditional distribution, $p(y_i|t_{0i})$, must be found and the maximum likelihood method can be applied (see Fuller, 1987, Ch 1.6.) For our purpose a simpler approach based on the results by Fedorov, 1974, see also Seber and Wild, 1989, Ch 10.9, appears to be sufficient.

Let us first assume that σ and δ are known and also that δ is “relatively” small. Assuming the existence of all the needed moments of the distribution under consideration and using Taylor’s expansion of (6) with respect to δ one can show:

$$E(y_i|t_{0i}, \theta) = \eta(t_{0i}, \theta) + \frac{1}{2}\delta^2\ddot{\eta}(t_{0i}, \theta) + o(\delta^2), \quad (7)$$

$$Var(y_i|t_{0i}, \theta) = \sigma^2 + \delta^2\dot{\eta}^2(t_{0i}, \theta) + o(\delta^2), \quad (8)$$

where E and Var mean “expectation” and “variance” correspondingly, $\dot{\eta}$ and $\ddot{\eta}$ mean the first and second derivatives of η with respect to t . From (8) it is easy to see that the “total” uncertainty in the observations have the tendency to increase in regions where the absolute value of the first derivative $\dot{\eta}(t, \theta)$ increases. This fact may explain the behavior of residuals in Fig 1a and b.

4. Estimation method

If one wants to stay within the realm of the least squares method then the estimator (the higher order terms are neglected):

$$\theta^* = \text{Arg} \min_{\theta \in \Omega} \sum_{i=1}^n \left[\frac{y_i - \eta(t_{0i}, \theta) - \frac{1}{2} \delta^2 \ddot{\eta}(t_{0i}, \theta)}{\sigma^2 + \delta^2 \dot{\eta}^2(t_{0i}, \theta)} \right], \quad (9)$$

where Ω is the set of admissible values of θ , looks very reasonable. However, in Fedorov, 1974, it was shown that generally the estimator defined by (7) is not consistent and that the following "iterated" least squares estimators should be used instead:

$$\hat{\theta} = \lim_{s \rightarrow \infty} \theta_s, \quad (10)$$

$$\theta_s = \text{Arg} \min_{\theta \in \Omega} \sum_{i=1}^n w_{si} [y_{si} - \eta(t_{0i}, \theta)]^2, \quad (11)$$

where

$$\begin{aligned} w_{si}^{-1} &= \sigma^2 + \delta^2 \dot{\eta}^2(t_{0i}, \theta_{s-1}), \\ y_{si} &= y_i - \frac{1}{2} \delta^2 \ddot{\eta}^2(t_{0i}, \theta_{s-1}). \end{aligned}$$

Discussions of similar approaches (but in different statistical settings) may be found, for instance, in Charles, Frome and Yu, 1976, and Jennrich and Moore, 1975. Estimation procedure (10) assumes that a standard nonlinear least squares algorithm with weights is used repeatedly without any intrinsic changes in the algorithm itself or corresponding software.

When there exists an opportunity to develop an independent software then the following version of (10) based on the Newton-Raphson algorithm (see, for instance, Seber and Wild, 1989) looks more attractive from the computational standpoint:

$$\hat{\theta} = \lim_{s \rightarrow \infty} \theta_s, \quad (12)$$

$$\theta_s = \theta_{s-1} + \zeta_s M_s^{-1} Y_s, \quad (13)$$

$$M_s = \sum_{i=1}^n w_{si} f_{si} f_{si}^T,$$

$$Y_s = \sum_{i=1}^n w_{si} y_{si} f_{si},$$

$$f_{si} = f(t_i, \theta_{s-1}), \quad f(t, \theta) = \frac{\partial \eta(t, \theta)}{\partial \theta}.$$

When $\zeta_s \equiv 1$, then 1 can be considered as the solution of the linearized with respect to θ version of (11). Evidently the estimators (10) and (12) are statistically identical if one manages to prove that limits of the sequences generated by (10) and (12) coincide. In practice this coincidence may be guaranteed by a proper choice of the sequence $\{\zeta_s\}$. For the simplest choice of $\{\zeta_s\}$ when $\zeta_s \equiv 1$ the following asymptotic result takes place (we do not pursue the mathematical rigor here).

For sufficiently smooth $\eta(x, \theta)$ as a function of x and θ (and, in particular, for the logistic function considered in this paper) the probability P_n that the limits for (10) and (12) coincide tends to one when the number of observation n tends to infinity, i.e.

$$\lim_{n \rightarrow \infty} P_n = 1. \quad (14)$$

In other words, for relatively small sample sizes there exists an opportunity that (10) and (12) may be different when $\zeta_s \equiv 1$, but for larger samples the chances of that are negligible.

In the standard least squares method enforcing the monotonicity of the sequence

$$\nu_s = \sum_{i=1}^n w_{si} (y_{si} - \eta(t_{0i}, \theta_{s-1}))^2 \quad (15)$$

through the proper choice of ζ_s (compare, for instance, with Seber and Wild, 1989) usually assures the convergence of an iterative procedure.

Enforced monotonicity in the considered case may cause estimator (12) to be inconsistent (compare with Fedorov, 1974). To explain this phenomenon the reader must notice that generally it can not be stated that

$$\min_{\theta \in \Omega} \sum_{i=1}^n w_{si} [y_{si} - \eta(t_{0i}, \theta)]^2 > \lim_{s \rightarrow \infty} \min_{\theta \in \Omega} \sum_{i=1}^n w_{si} [y_{si} - \eta_i(t_{0i}, \theta)]^2 .$$

Actually in practice the inverse can be observed frequently. Estimator (10) is consistent and asymptotically normally distributed in the framework of approximation (7), (8), see Fedorov, 1974. Therefore, for sufficiently large number of observations the variance-covariance of the estimated parameters approximately equals:

$$Cov(\theta^*) \simeq \sum_{i=1}^n w_i f(t_i, \theta^*) f^T(t_i, \theta^*) \quad (16)$$

where $w_i = \lim_{s \rightarrow \infty} w_{si}$, and $f(t, \theta)$ is defined in the comments to (12), (13).

Matrix (16) may be computed inside of a standard nonlinear least squares routine used to solve (11). One can notice also that

$$Cov(\theta^*) = \lim_{s \rightarrow \infty} M_s,$$

where the matrix M_s is defined in comments to (12), (13).

So far, we have assumed that the standard errors σ and δ are given. Of course, in most practical situations those parameters have to be estimated. Details on how to modify (10) to estimate σ and δ are discussed in Section 5.

For the experiments analyzed in DHN, 1990, the needed standard errors can be estimated independently of (10). As a matter of fact, there are a number of points with repeated observations and some of them are allocated where the response function (see Fig. 1-4) achieves its upper limit and correspondingly the first derivative of response function vanishes. Therefore,

$$\text{Var}(y_i|t_{0i}, \theta) \simeq \sigma^2$$

and for the first stage of this study the simple estimator

$$\hat{\sigma}^2 = \frac{1}{n_i - 1} \sum_{j=1}^{n_i} (y_{ij} - \bar{y}_i)^2,$$

$$\bar{y}_i = \frac{1}{n_i} \sum_{j=1}^{n_i} y_{ij}$$

was used.

The parameter δ was chosen as the accuracy of achieving the desired temperature. With a view towards analyzing a boundary situation, δ was selected to be 15C. This selection was chosen following a review of Nanstad et al. (1990), which considered the influence of various thermal conditioning on media on temperature changes experienced by the test specimen during Charpy impact testing. The value chosen and used for this study is not a value normally observed or expected. The estimation procedures (8) can be easily embedded in the standard reweighted least squares algorithm. In particular we have used the nonlinear least squares procedure with reweighting from SAS, 1990. In (8) the set Ω was defined by inequalities: $\theta_1 \geq 0$, $\theta_2 \geq 0$, $\theta_3 \geq 0$. Fig. 5-8 give a graphical presentation of the results, while Table 1 complements them with the parameter estimates and their asymptotic standard errors.

Everywhere the iterations were started with $w_{0i}^{-1} = \hat{\sigma}^2$ and $y_{0i} = y_i$. Therefore the results of the first iteration coincide with the least squares estimates (dotted lines). One can see that the iterated estimation does not change the response

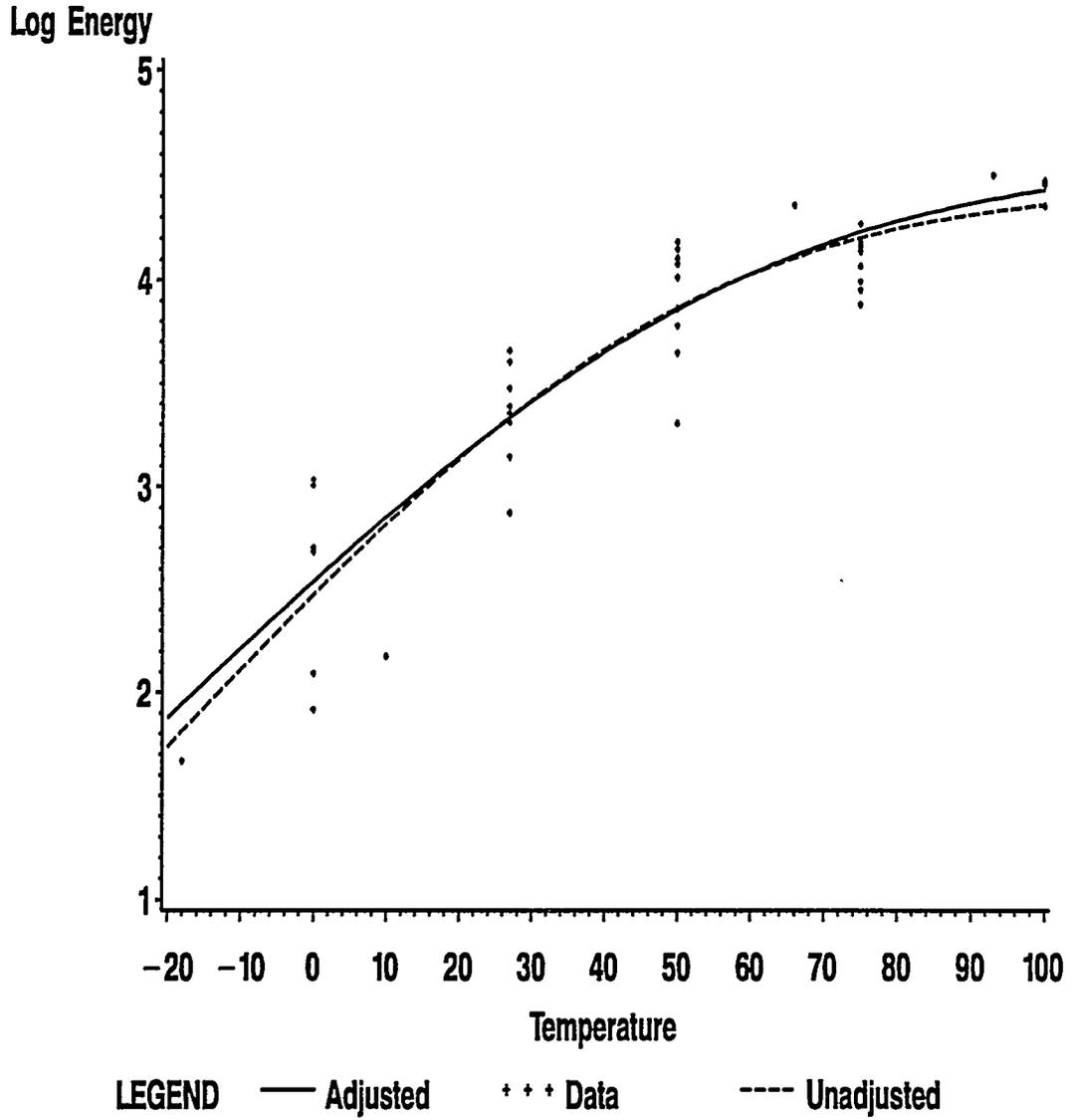


Figure 5: Response function for the 72W irradiated data. Calculated with and without adjusting for error in temperature.

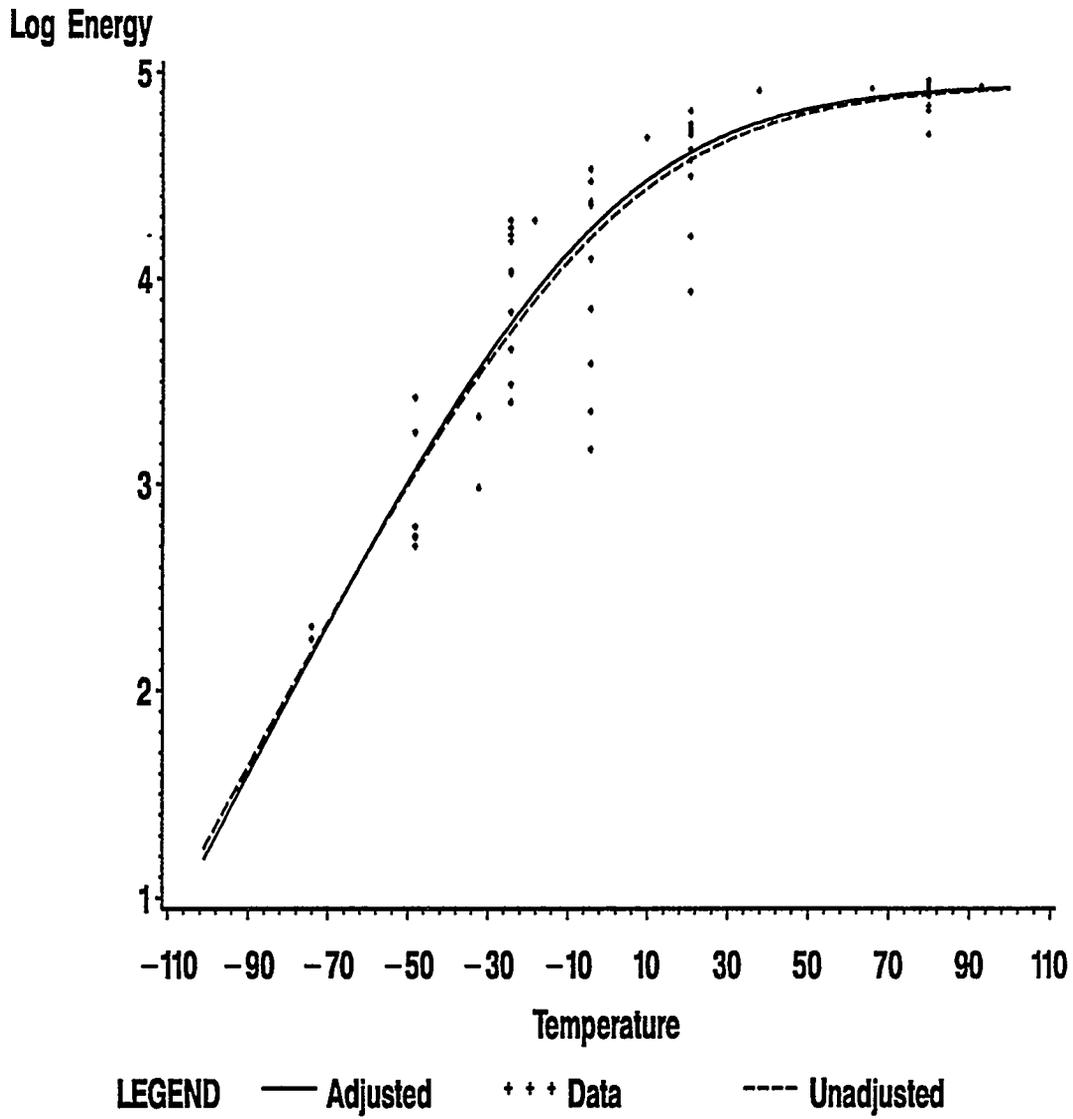


Figure 6: Response function for the 72W unirradiated data. Calculated with and without adjusting for error in temperature.

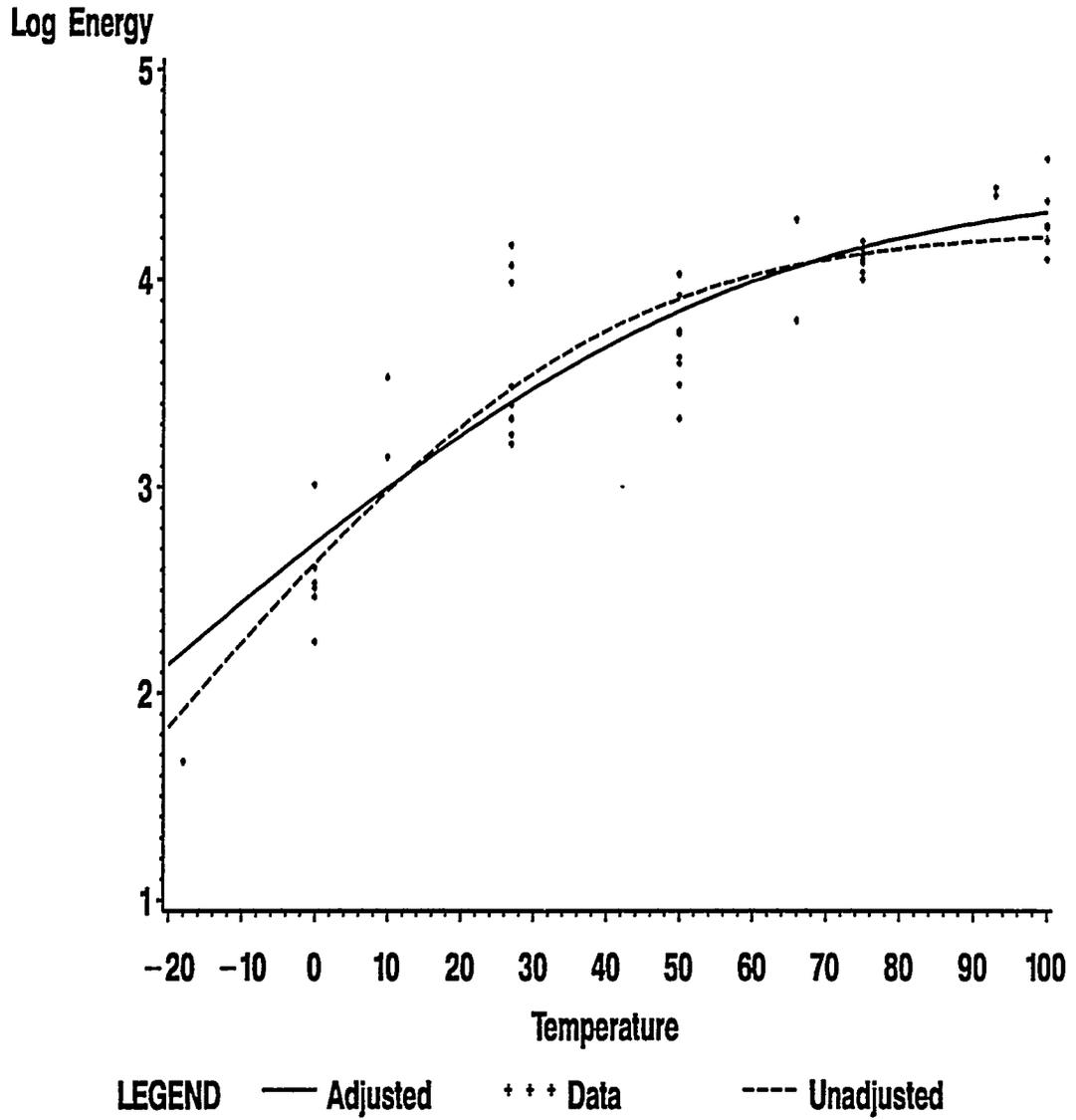


Figure 7: Response function for the ⁷³W irradiated data. Calculated with and without adjusting for error in temperature.

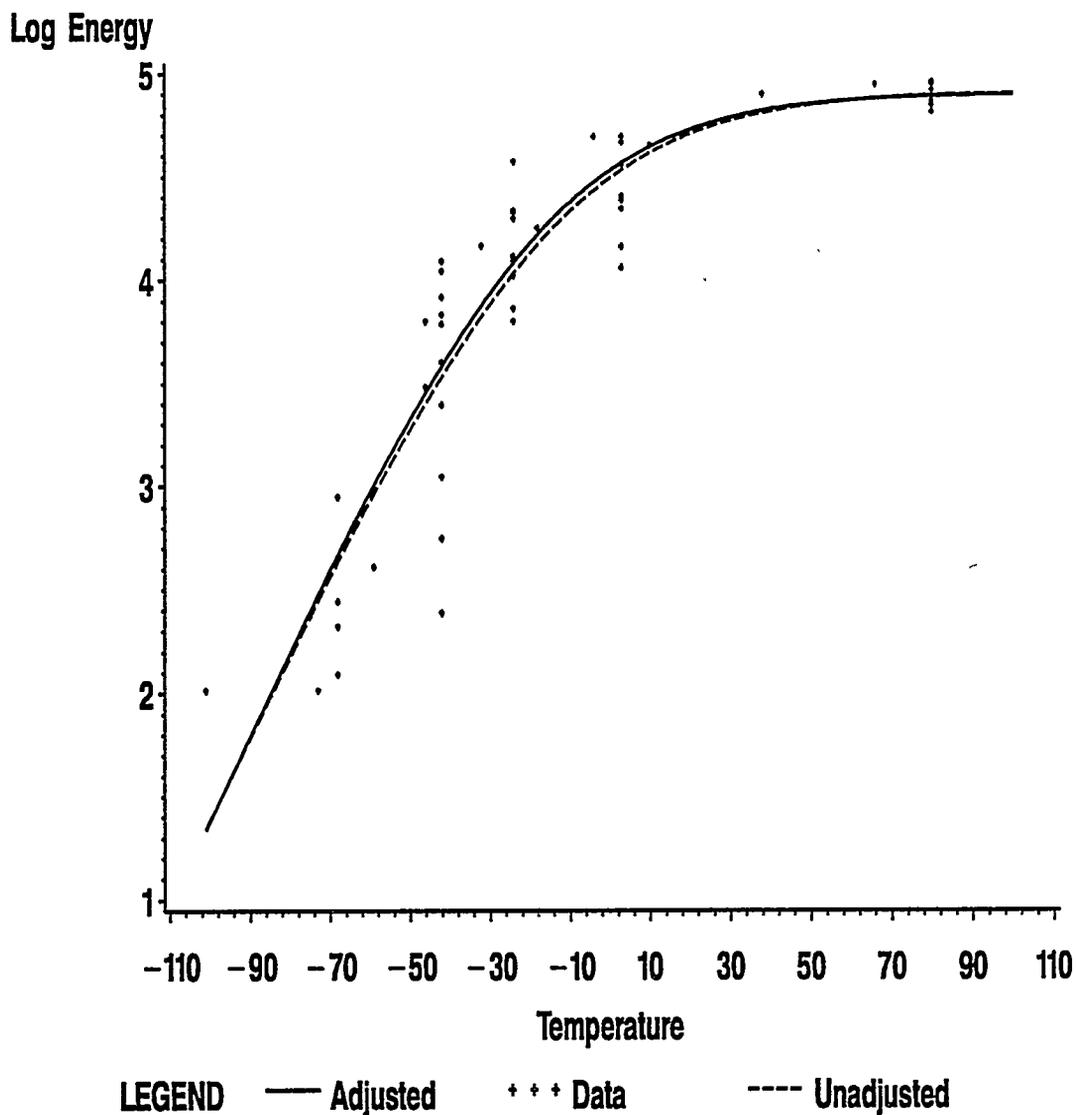


Figure 8: Response function for the 73W unirradiated data. Calculated with and without adjusting for error in temperature.

Table 1: Parameter estimates and their standard errors.

<i>Dataset</i>	<i>Parameter</i>	<i>Standard LSE</i>		<i>Iterated LSE</i>	
		<i>Estimate</i>	<i>Standard Error</i>	<i>Estimate</i>	<i>Standard Error</i>
72I	u	4.45	0.054	4.58	0.137
	θ_3	0.04	.002	0.04	0.006
	θ_4	44.75	1.647	52.17	5.023
72U	u	4.93	0.033	4.93	0.040
	θ_3	0.04	0.001	0.04	0.004
	θ_4	-1.57	2.265	-3.70	4.274
73I	u	4.23	0.035	4.46	0.116
	θ_3	0.05	0.002	0.03	0.006
	θ_4	29.94	1.217	45.55	4.766
73U	u	4.90	0.030	4.89	0.033
	θ_3	0.04	0.001	0.04	0.004
	θ_4	-15.98	2.113	-18.60	4.844

function too much. As expected the most significant corrections appear in areas where the second derivatives are relatively large. Evidently the least squares estimation of the response function leads to a negative bias when the second derivative is positive and to a positive bias when it is negative, compare with (7).

The values of the parameters estimates do not change significantly going from the unweighted to the weighted/iterated estimates. We would like to mention here that estimates of θ_1 were not significantly different from 0 in all the cases and therefore this parameter was excluded from further analysis. Since θ_1 was excluded then we may use a more convenient version of (3):

$$\eta(t_i, \theta) = u - \ln(1 + \exp(\theta_3(t - \theta_4))) \quad (17)$$

Note in this parameterization that u has taken the place of $\ln(\theta_2)$ and u is given in Table 1.

The most noticeable changes in the results occurred to the “accuracy” measures of the estimators. We can see than almost twofold increase in the values of the standard error estimates. Actually only the numbers reported for the last

iteration can be considered as the asymptotic variances of the estimators. The values, which appear at the first iteration and are usually reported as the variances in the least squares method may support false expectations of a practitioner about the quality of the standard least squares estimates. In reality, the variances of the least squares estimates will be greater than the values reported in Table 1. Moreover the estimators of the parameters are not consistent due to the bias generated by (7).

We plotted the standardized residuals (see Fig. 9-16)

$$r_{si} = w_{si} [y_{si} - \eta(t_i, \hat{\theta}_s)]$$

for $\delta = 0$ and $\delta^* = 15$, when the procedure had converged. Since the residuals fall in a horizontal band across all temperatures, one can see that for δ^* there is no more heteroscedasticity which is so prominent with $\delta = 0$ (for the standard least squares estimators). In other words, based on a $\delta = 15$ (considered a bounding value) and neglecting metallurgical factors, the methodology developed herein may explain the heteroscedasticity observed in Charpy impact populations such as those in DHN, 1990. We would like to emphasize at this point that perhaps other explanations for the heteroscedasticity of the original regression model can be found and they can be related to the change of the response function or the type of assumed distribution on the errors ϵ_i . Our analysis tried to remove the heteroscedasticity within the framework of the previously accepted "physical" model.

5. Unknown σ and δ

Iterative procedure (10) together with (11) leads to the results which have explained the appearance of heteroscedasticity in DHN, 1990. Nevertheless, two parameters: σ and δ were evaluated in a rather ad hoc fashion. The whole estimation procedure will become better formalized if the iterative estimation in-

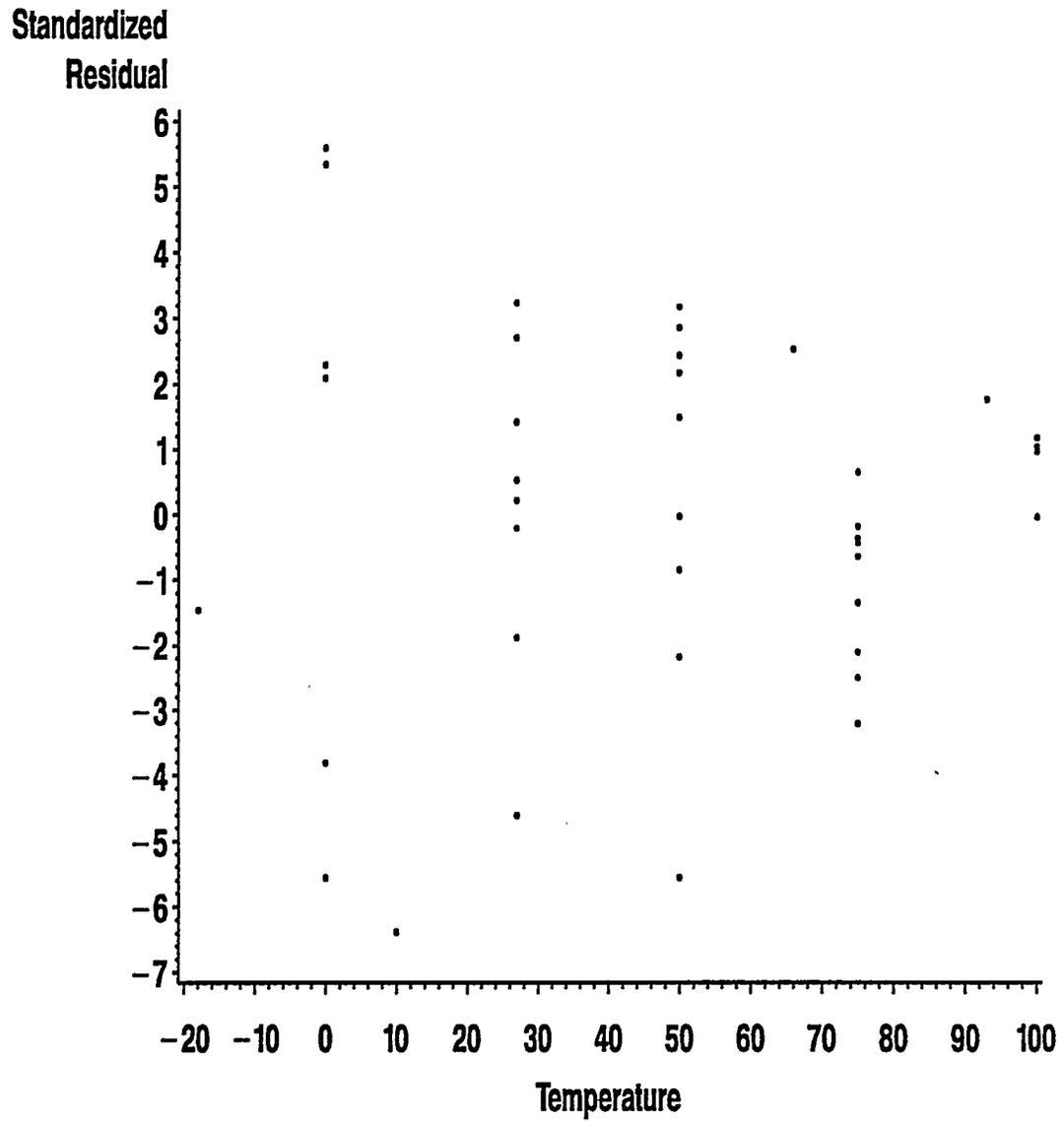


Figure 9: Standardized residuals for the 72W irradiated data. Not adjusted for error in temperature.

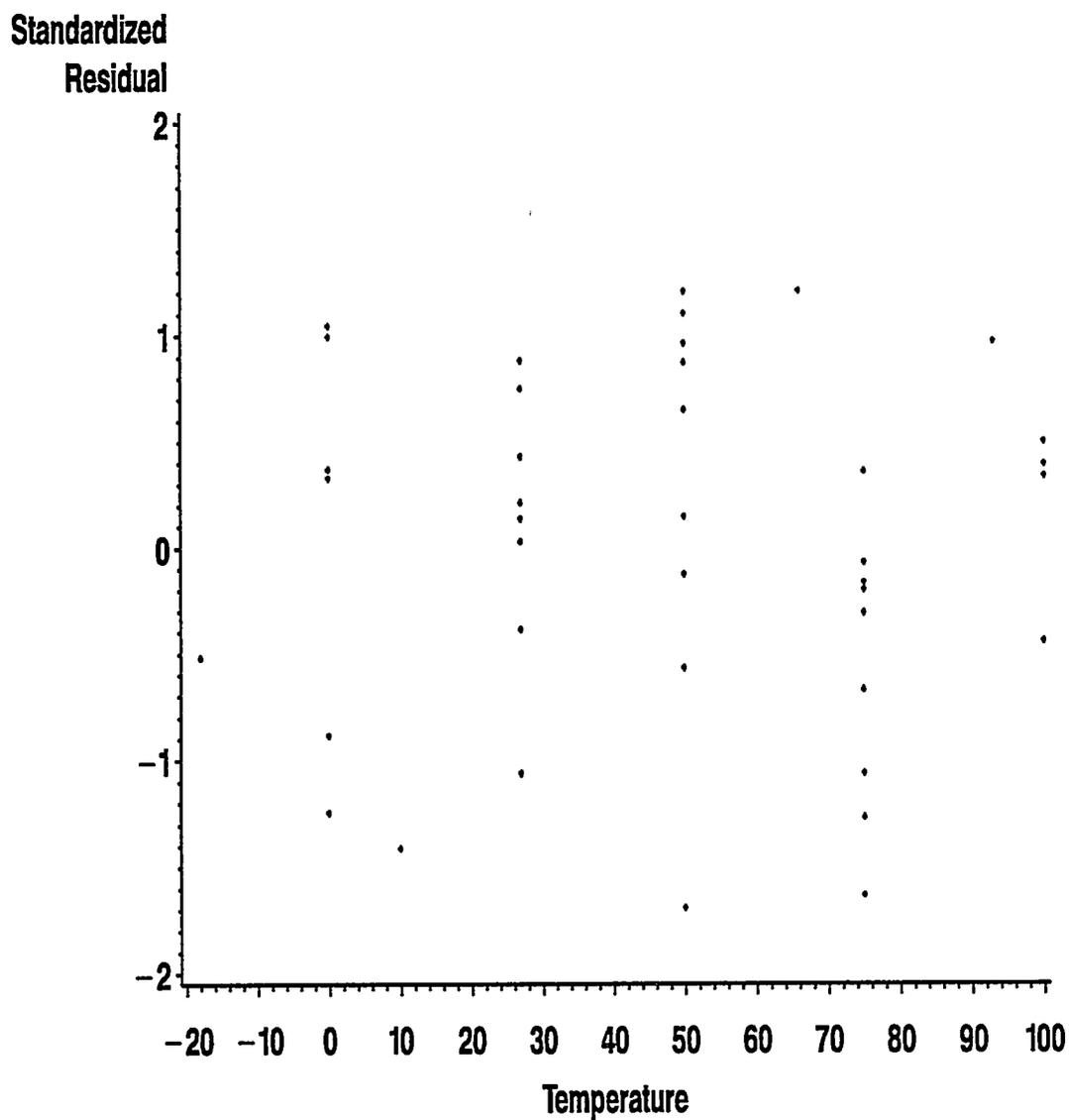


Figure 10: Standardized residuals for the 72W irradiated data. Adjusting for error in temperature.

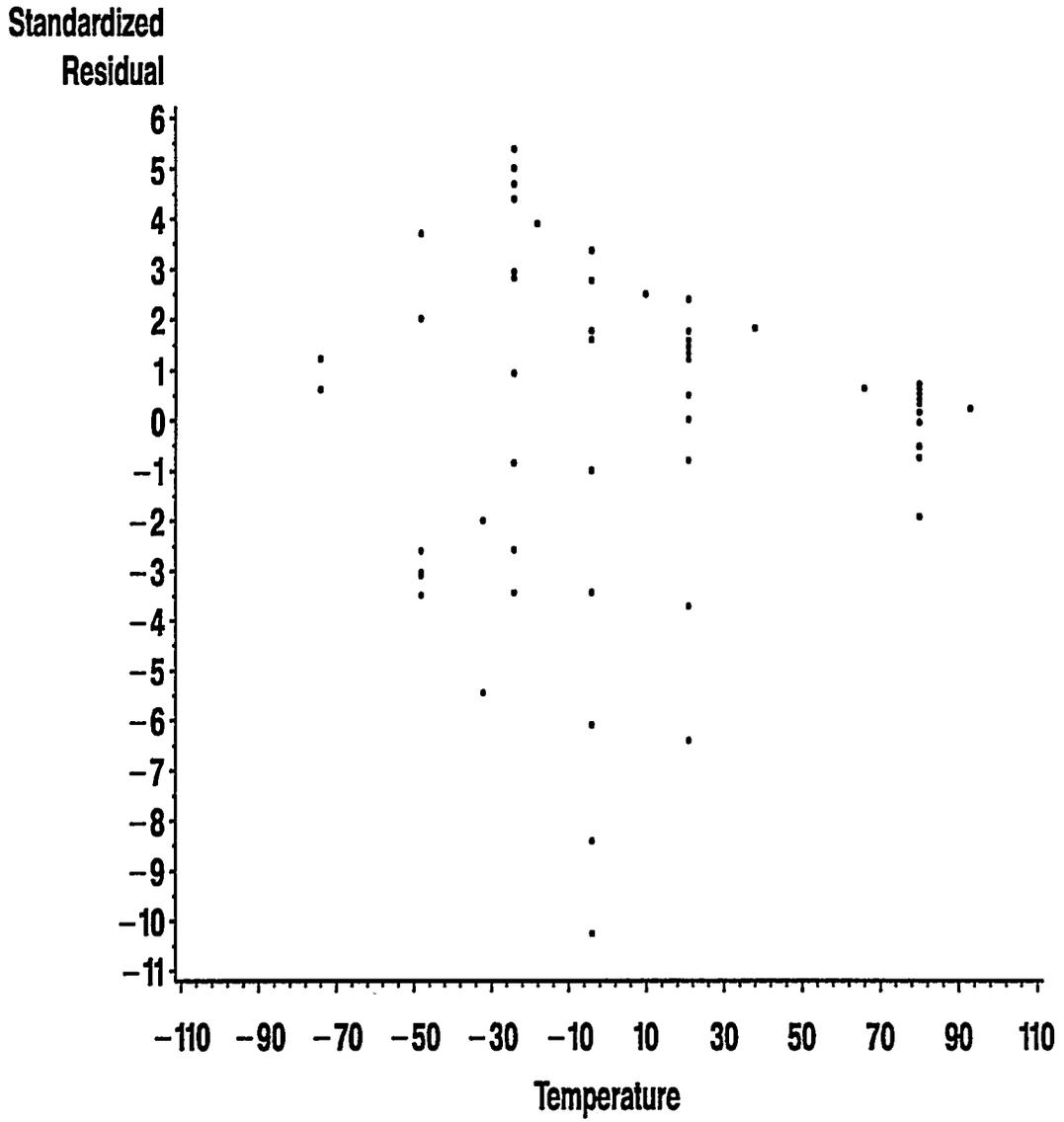


Figure 11: Standardized residuals for the 72W unirradiated data. Not adjusted for error in temperature.

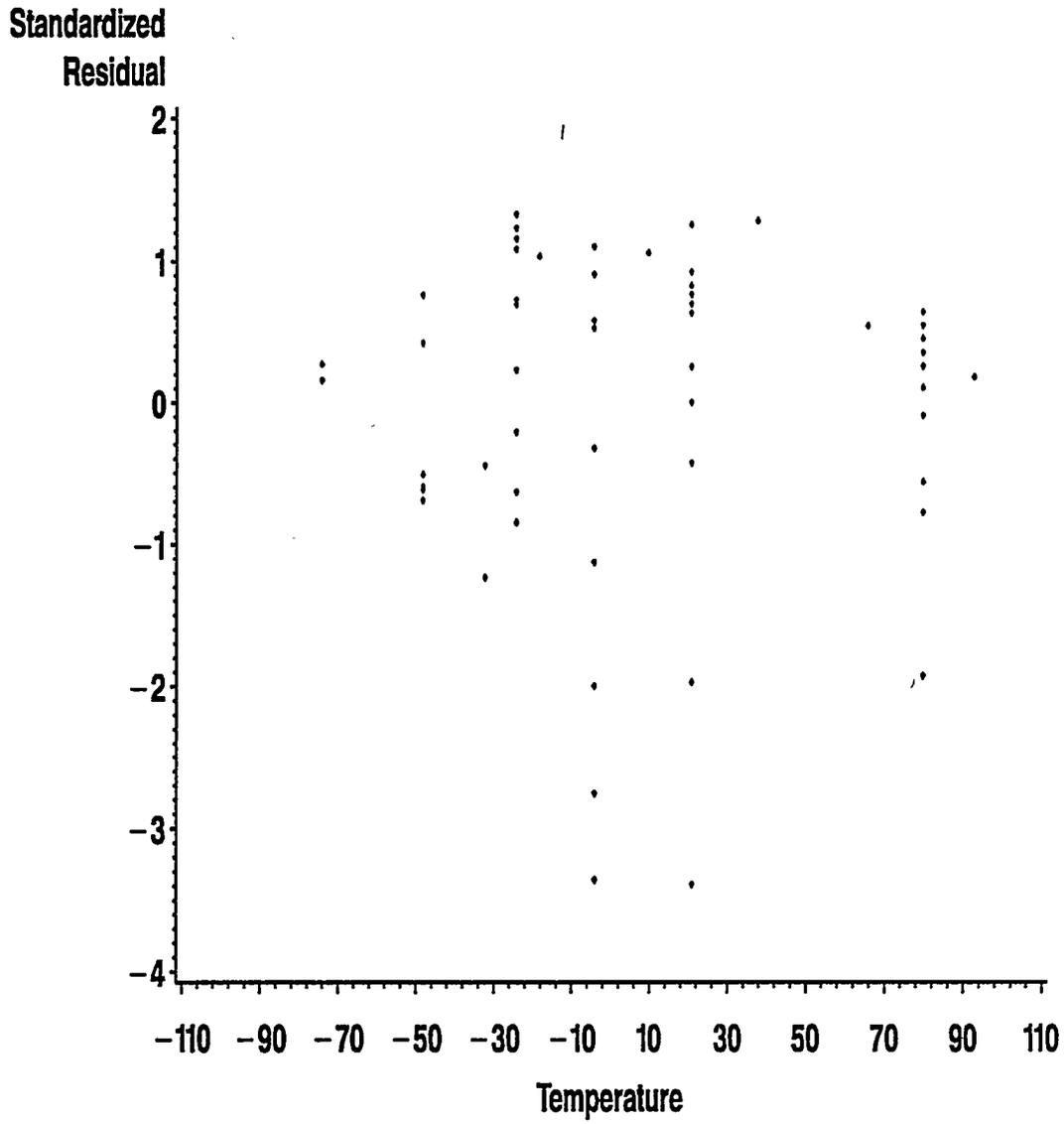


Figure 12: Standardized residuals for the 72W unirradiated data. Adjusting for error in temperature.

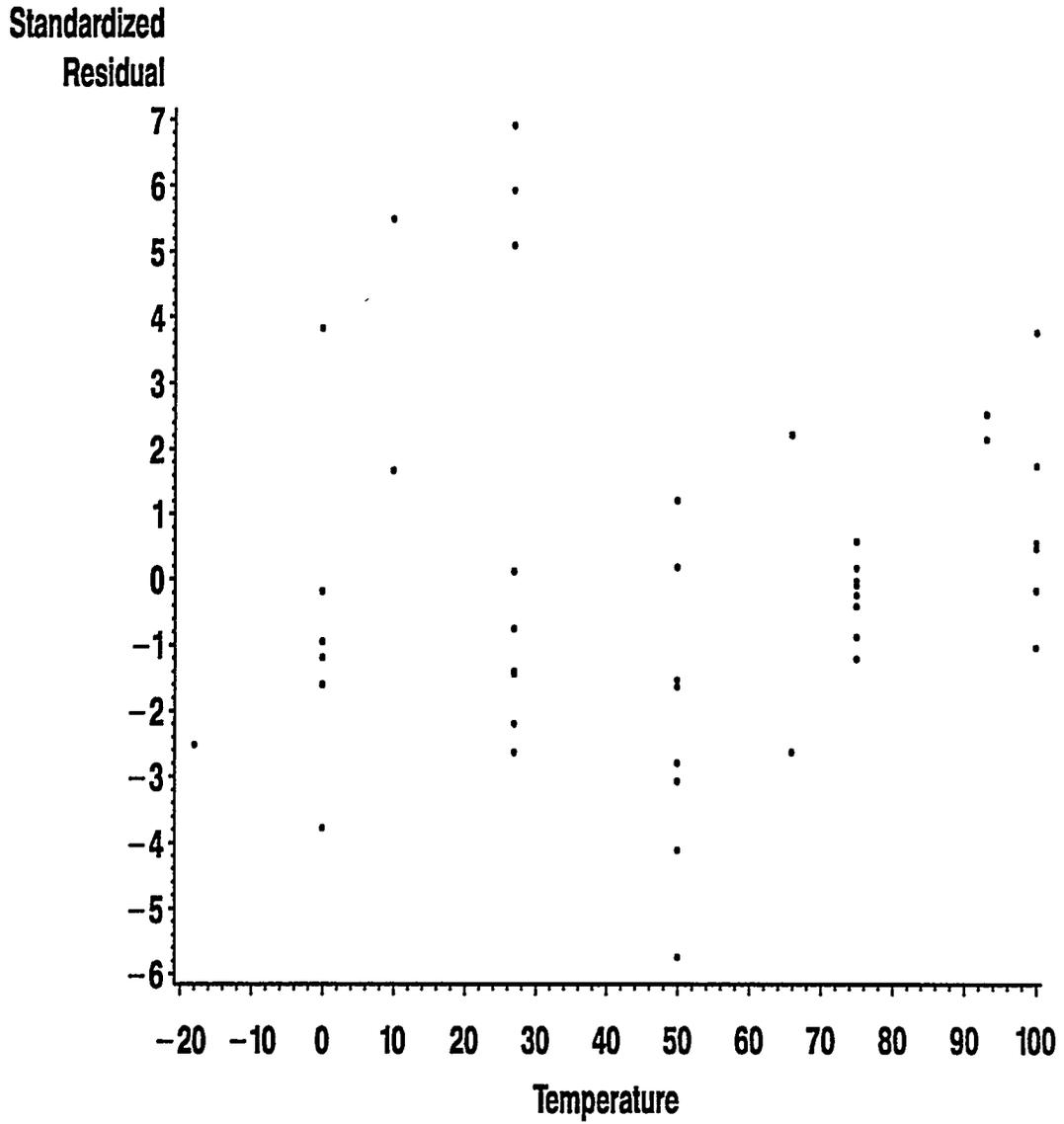


Figure 13: Standardized residuals for the 73W irradiated data. Not adjusted for error in temperature.

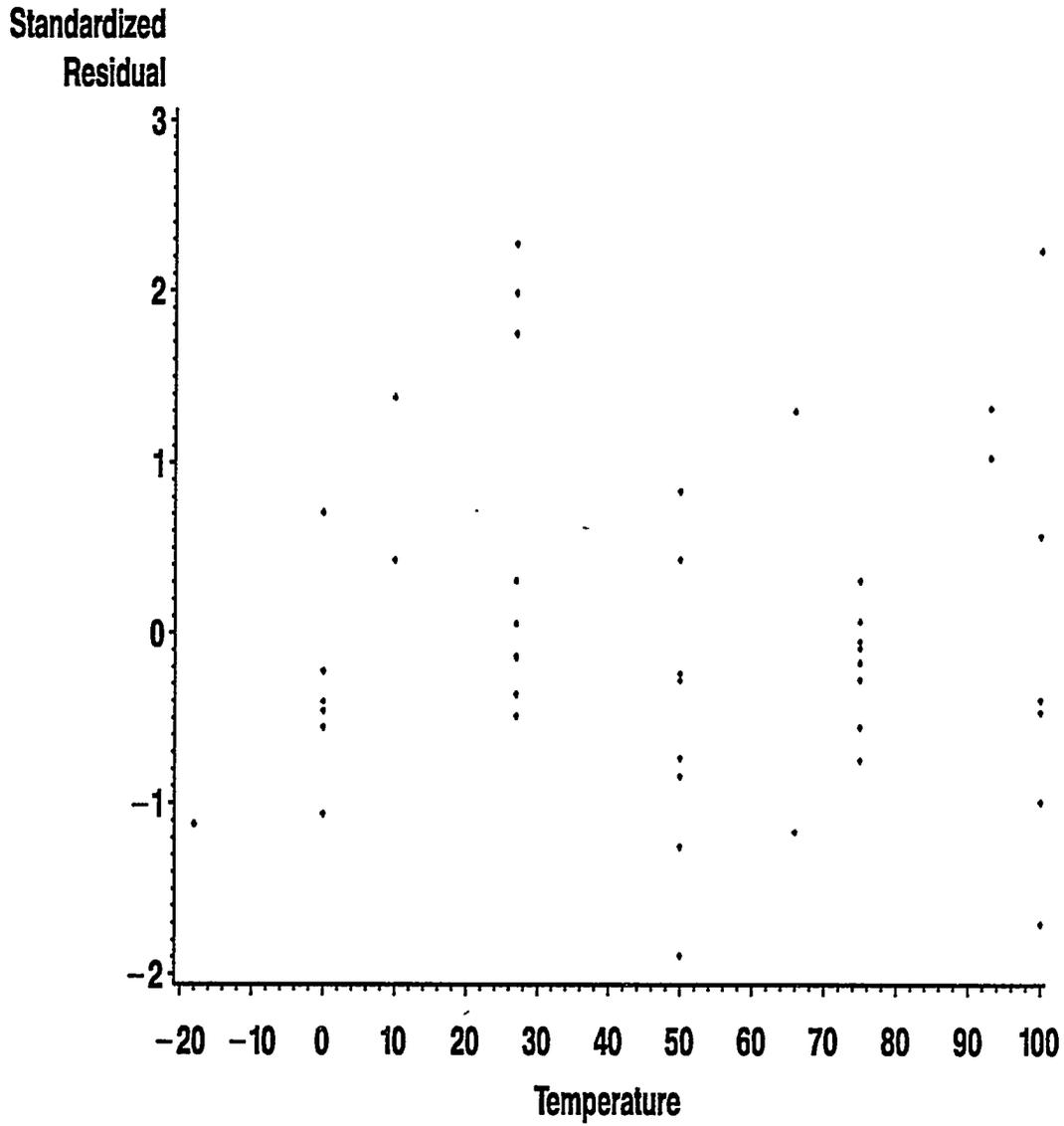


Figure 14: Standardized residuals for the 73W irradiated data. Adjusting for error in temperature.

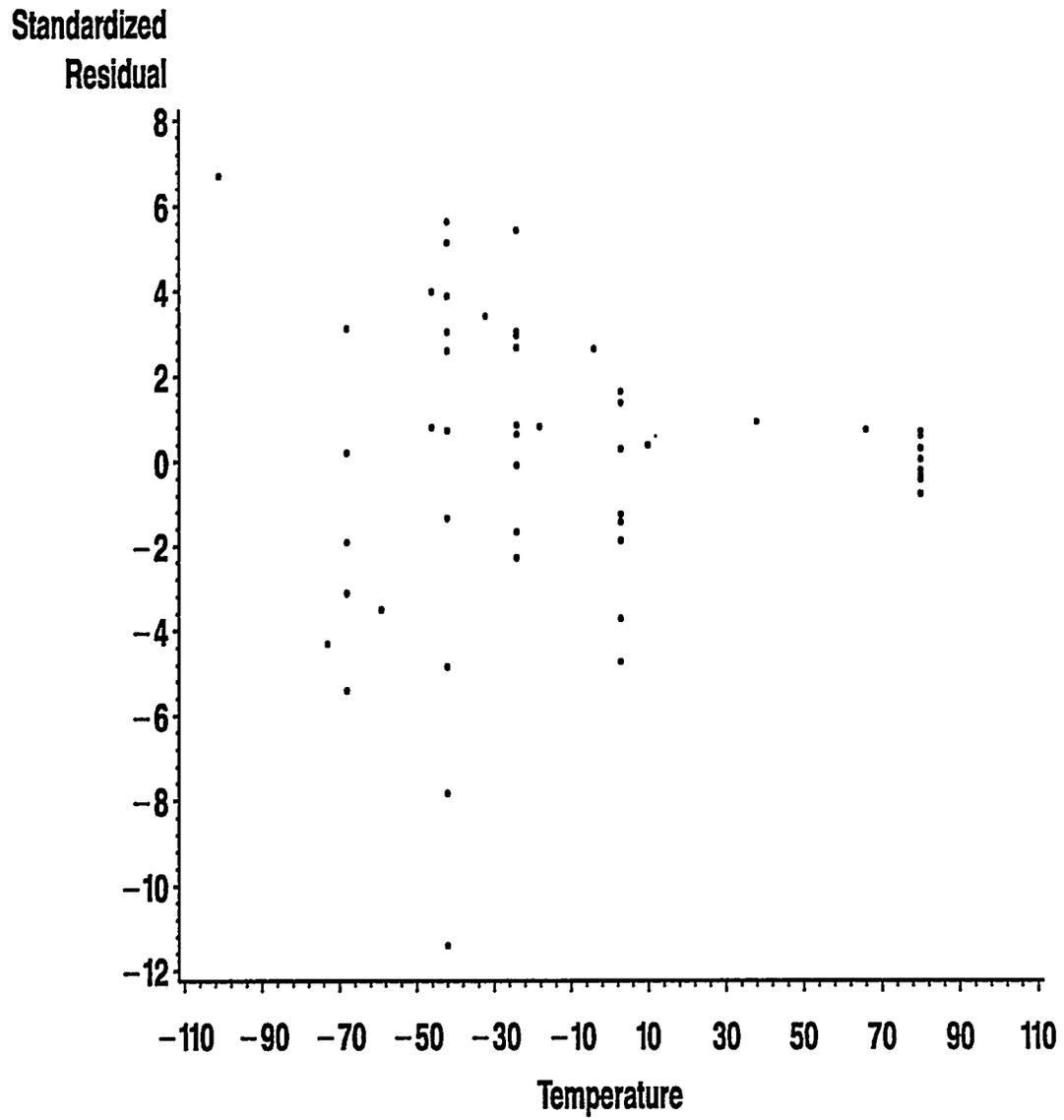


Figure 15: Standardized residuals for 73W unirradiated data. Not adjusted for error in temperature.

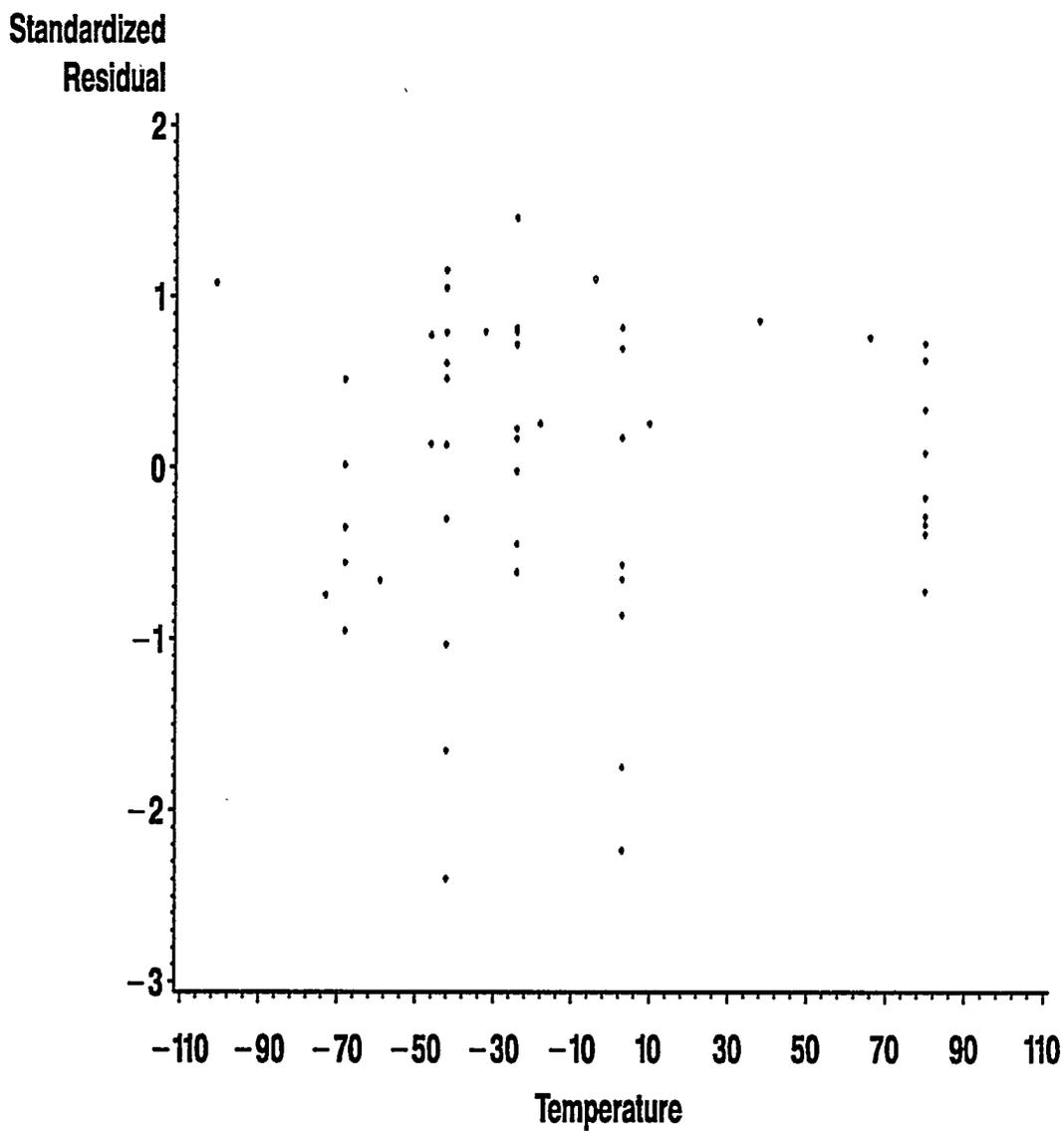


Figure 16: Standardized residuals for the 73W unirradiated data. Adjusting for error in temperature.

cludes σ and δ . The corresponding statistical technique for that was considered in a number of publications, and probably the closest regression problems were analyzed in Malyutov, 1982.

Application of the ideas of the iterated estimation to the extended set of unknown parameters: θ, σ and δ yields the following iterative estimators:

$$(\hat{\theta}, \hat{\sigma}, \hat{\delta}) = \lim_{s \rightarrow \infty} (\hat{\theta}_s, \hat{\sigma}_s, \hat{\delta}_s), \quad (18)$$

$$\theta_s = \text{Arg} \min_{\theta \in \Omega} \sum_{i=1}^n w_{si} [y_{si} - \eta(t_{0i}, \theta)]^2, \quad (19)$$

$$(\delta_s, \sigma_s) = \text{Arg} \min_{\delta, \sigma} \sum_{i=1}^n w_{si}^2 [u_{si} - v_s(t_{0i}, \sigma, \delta)]^2, \quad (20)$$

where

$$\begin{aligned} w_{si}^{-1} &= \sigma_{s-1}^2 + \delta_{s-1}^2 \dot{\eta}^2(t_{0i}, \theta_{s-1}), \\ v_s(t_{0i}, \sigma, \delta) &= \sigma^2 + \delta^2 \dot{\eta}^2(t_{0i}, \theta_s), \\ u_{si} &= [y_{si} - \eta(t_{0i}, \theta_s)]^2. \end{aligned}$$

From the computational point of view optimization problems (19) and (20) are standard least squares problems. Moreover, (20) is linear. It is evident that (18) - (20) may be replaced by the procedure similar to (12), (13). In our calculations we have used this modified version. We repeated the analysis with the enhanced procedure. Amazingly our initial guesses about σ and δ were very close to the estimated results. Consequently all the results are very close and we do not report the latter ones here.

6. Conclusions

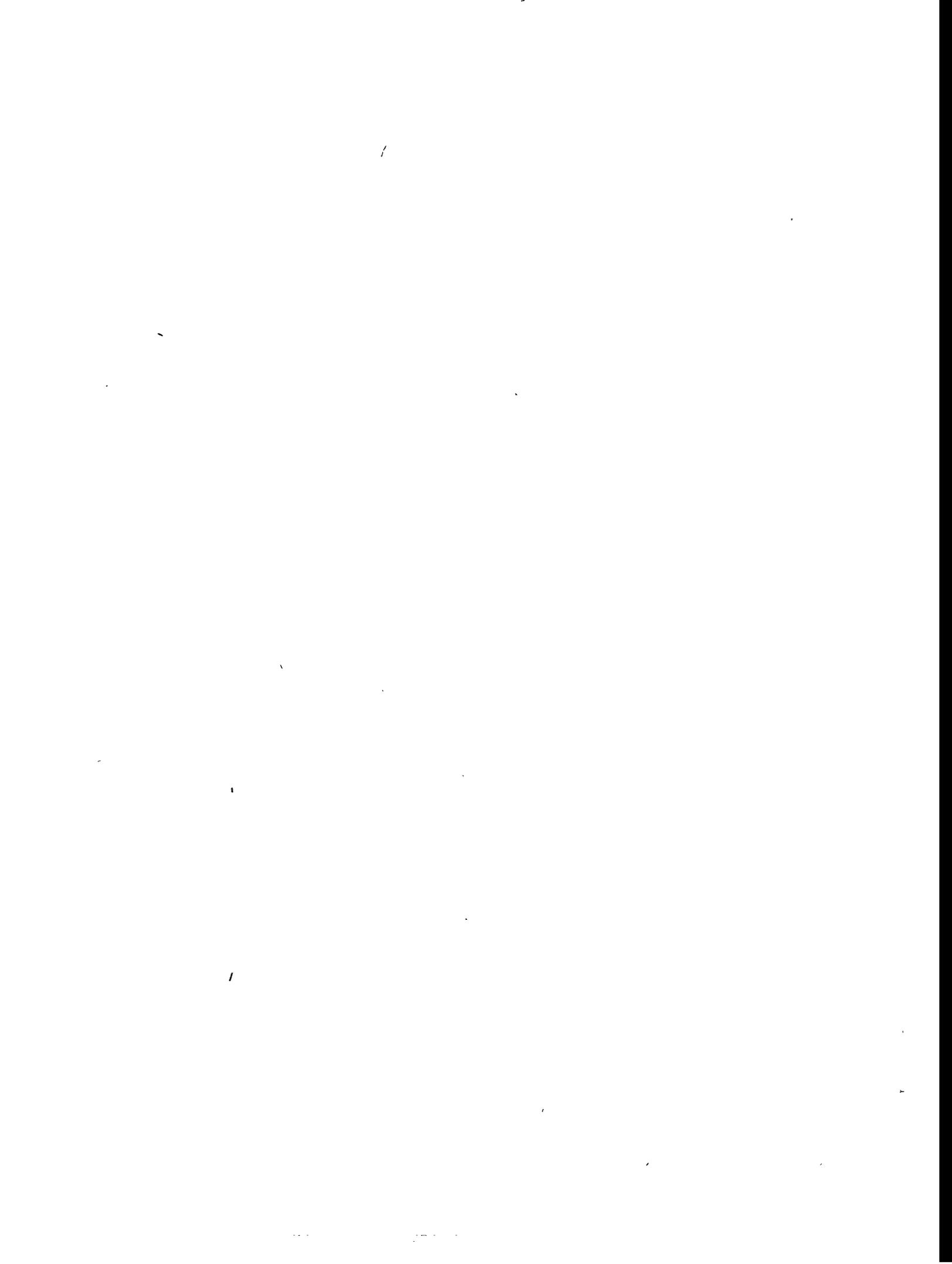
1. Heteroscedasticity of residuals may be a consequence of not being able

to control independent variables as well as we think. The effect is more pronounced in areas for which the response function has greater slope.

2. Results of using the usual least squares estimators make us think that we are doing better than we are, i.e. the standardly reported "accuracy" measures are not true. When the curvature of the response function is significant the usual least squares estimators are not consistent and may show a noticeable bias even for moderate sample sizes.
3. Methods are proposed to estimate all parameters including variances of errors both in response and predictors. These methods allow one to develop algorithms based on software available through commercial packages (SAS, for instance was used in this study).

7. References

- [1] Downing, D. J., Haggag, F. M. and Nanstad, R. K. (1990), "Estimating Charpy Transition Temperature Shift Using Weibull Analysis", *Int. J. Pres. Ves. and Piping*, **44**, 241-254.
- [2] Charnes, A., Frome, E. L., and Yu, P. L. (1976), "The Equivalence of Least Squares and Maximum Likelihood Estimates in the Exponential Family", *JASA*, **71**, 216-222.
- [3] Fedorov, V. V. (1974), "Regression Problems with controllable Variables Subject to Error", *Biometrika*, **61**, 49-56.
- [4] Helm, J. L. (1995), "The Interpretation of Charpy Impact Test Data Using Hyper-Logistic Fitting Functions", in *Effects of Radiation on Materials: 17th International symposium*, ASTM STP 12xx, Gelles, D. S., Nanstad, R. K., Kumar, A. S., and Little, E. H., Editors, American Society for Testing and Materials, Philadelphia.
- [5] Jennrich, R. I., and Moore, R. H. (1975) "Maximum Likelihood Estimation by Means of Nonlinear Least Squares", *Am. Stat. Assoc., Proc. Statist. Comput. Section*, 57-65.
- [6] Malyutov, M. B. (1982), "Asymptotical Properties and Applications of the IRGINA-estimator of Parameters of Generalized Regression Model", in *Stochastic Processes and Applications*, Moscow, pp. 144-158.
- [7] Seber, G. A. F., and Wild, C. T. (1989), *Nonlinear Regression*, Wiley, New York.
- [8] Fuller, W. A. (1987), *Measurement Error Models*, Wiley, New York.
- [9] SAS (1990) Version 6, Fourth Edition, Cary, North Carolina.



ORNL/TM -13140

INTERNAL DISTRIBUTION

- | | |
|-----------------------|---------------------------------|
| 1. D. J. Alexander | 33. R. K. Nanstad |
| 2. C. K. Bayne | 34. C. E. Oliver |
| 3. W. R. Corwin | 35-39. G. Ostrouchov |
| 4. D. F. Craig | 40-44. S. A. Raby |
| 5. T. S. Darland | 45. R. L. Schmoyer |
| 6. C. S. Daw | 46. R. F. Sincovec |
| 7-11. D. J. Downing | 47. P. T. Singley |
| 12. L. J. Gray | 48. M. A. Sokolov |
| 13. H. W. Hayden, Jr. | 49. D. A. Wolf |
| 14. S. K. Iskander | 50. K-25 Applied Tech. Library |
| 15. P. Kanciruk | 51. Y-12 Technical Library |
| 16-20. W. F. Lawkins | 52. Laboratory Records - RC |
| 21-25. M. R. Leuze | 53-54. Laboratory Records Dept. |
| 26. D. E. McCabe | 55. Central Research Library |
| 27. J. G. Merkle | 56. M&C Records Office |
| 28-32. M. D. Morris | 57. ORNL Patent Office |

EXTERNAL DISTRIBUTION

58. Dr. Richard Beckman, Statistics Group A1, Los Alamos National Laboratory, MS F600, Los Alamos, NM 87545
59. C. Bolton, Nuclear Electric, Berkeley Nuclear Laboratories, Berkeley, Gloucestershire, GL139PB, United Kingdom
60. Prof. Dennis Cox, Department of Statistics, Rice University, Houston, TX 77251-1892
61. E. Eason, Modeling and Computing Services, Suite 105, 6400 Lookout Road, Boulder, CO 80301
62. Prof. Sherwood Ebey, Department of Mathematics, University of the South, Sevanee, TN 37375
63. A. Fabry, Central Institute for Nuclear Studies, SCK-CEN, Boeretang 200, B - 2400 Mol, Belgium
64. Dr. David Hall, Statistics, Systems Department, Pacific Northwest Laboratory, P.O. Box 999, Richland, WA 99352
65. Prof. J. A. Hartigan, Department of Statistics, Yale University, Box 2179 - Yale Station, New Haven, CT 06520
66. J. Helm, Department of Applied Physics, 202 Mudd Building, Columbia University, New York, NY 10027

67. Dr. Dan Hitchcock, Office of Scientific Computing, ER-7, Applied Mathematical Sciences, Office of Energy Research, U. S. Department of Energy, Washington, DC 20585
68. Dr. Fred Howes, Office of Scientific Computing, ER-7, Applied Mathematical Sciences, Office of Energy Research, U. S. Department of Energy, Washington, DC 20585
69. Prof. Mark Johnson, Department of Statistics, University of Central Florida, Orlando, FL 32816-0370
70. Dr. A. M. Liebetrau, Computational Sciences Department, Battelle-Northwest, P. O. Box 999, Richland, WA 99352
71. Dr. J. Lijengren, Pacific Northwest Laboratories, P. O. Box 999, Richland, WA 99352
72. R. McElroy, AEA Technology, Harwell Laboratory, Oxfordshire OX11 0RA, United Kingdom
73. Dr. Michael McKay, Statistics Group A1, Los Alamos National Laboratory, MS F600, Los Alamos, NM 87545
74. Prof. Lisa Moore, Institute of Statistics and Decision Sciences, Duke University, Durham, NC 27706
75. Dr. David Nelson, Director of Scientific Computing, ER-7, Applied Mathematical Sciences, Office of Energy Research, U. S. Department of Energy, Washington, DC 20585
76. G. R. Odette, Department of Chemical and Nuclear Engineering, Ward Memorial Drive, University of California, Santa Barbara, CA 93106
77. Dr. David J. Pack, CSR Institute, 8889 Bourgade AV, Lenaxa, KS 66219
78. Mr. Brent Pulsipher, Computational Sciences Department, Battelle Northwest, P. O. Box 999, K1-86, Richland, WA 99352
79. Dr. Jerome Sacks, NISS, P. O. Box 14162, Research Triangle Park, NC 27709-4162
80. Prof. A. F. Smith, Department of Mathematics, University of Nottingham, University Park, Nottingham NG7 2RD, England
81. Dr. Alan Solomon, P. O. Box 227, Omer 84965, Israel
82. Dr. Daniel L. Solomon, Department of Statistics, North Carolina State University, P. O. Box 5457, Raleigh, NC 27650
83. A. Taboada, Mail Stop T0-E10, Division of Engineering, U.S. Nuclear Regulatory Commission, Washington, DC 20555
84. M. G. Vassilaros, MS T10-E10, Division of Engineering, U. S. Nuclear Regulatory Commission, Washington, DC 20555
85. K. Wallin, Senior Consultant, Technical Research Centre of Finland, Kemistintie 3, Box 2G, Espoo 02750, Finland
86. Prof. William Welch, Department of Statistics and Actuarial Science, University of Waterloo, Waterloo, Ontario N2L 3G1, Canada

87. Prof. Y. C. Yao, Department of Statistics, Colorado State University, Fort Collins, CO 80523
88. Prof. Don Ylvisaker, Department of Mathematics, University of California, Los Angeles, CA 90024
89. Office of Assistant Manager for Energy Research and Development, Department of Energy, Oak Ridge Operations Office, P. O. Box 2001, Oak Ridge, TN 37831-8600
90. Office of Scientific and Technical Information, P. O. Box 62, Oak Ridge, TN 37830

