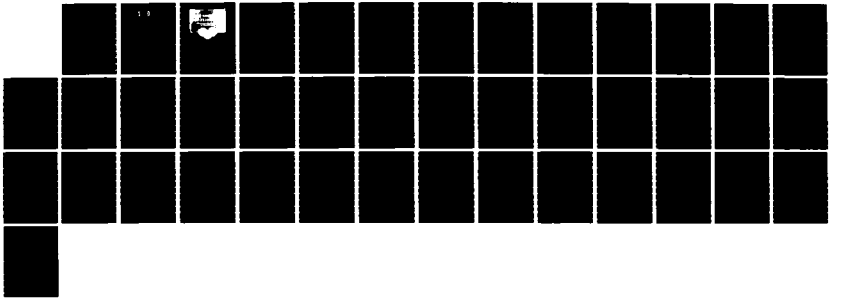
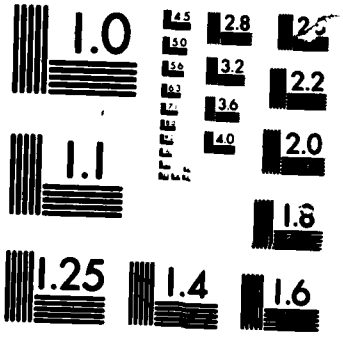
AD-A160 487 ESTIMATING WEIGHTS IN HETEROSCEDASTIC REGRESSION MODELS 1/1
BY APPLYING LEAST..(U) NORTH CAROLINA UNIV AT CHAPEL
HILL DEPT OF STATISTICS R J CARROLL ET AL. OCT 85
UNCLASSIFIED MIMEO- SER-1584 AFOSR-TR-86-0318 F/G 12/1 NL





MICROCOPY CHART

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE

(2)

REPORT DOCUMENTATION PAGE

1a. REPORT SECURITY CLASSIFICATION: UNCLASSIFIED	1b. RESTRICTIVE MARKINGS
2a. SECURITY CLASSIFICATION AUTH[illegible]	3. DISTRIBUTION/AVAILABILITY OF REPORT: Unlimited Approved for public release; distribution unlimited.
2b. DECLASSIFICATION/DOWNGRADING [illegible]EDULE	
4. PERFORMING ORGANIZATION REPORT NUMBER(S): Mimeo Series #1584	5. MONITORING ORGANIZATION REPORT NUMBER(S): AFOSR-TR- 86-0318
6a. NAME OF PERFORMING ORGANIZATION: Univ. of North Carolina; 6b. OFFICE SYMBOL (If applicable):	7a. NAME OF MONITORING ORGANIZATION: Air Force Office of Scientific Research
6c. ADDRESS (City, State and ZIP Code): Dept. of Statistics, Phillips Hall, 039A Chapel Hill, NC 27514	7b. ADDRESS (City, State and ZIP Code):
8a. NAME OF FUNDING/SPONSORING ORGANIZATION: AFOSR; 8b. OFFICE SYMBOL (If applicable):	9. PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER: F49620 82 C 0009
8c. ADDRESS (City, State and ZIP Code): Bolling Air Force Base, Washington, DC 20332	10. SOURCE OF FUNDING NOS. — PROGRAM ELEMENT NO.: 61102F; PROJECT NO.: 2304; TASK NO.: A5; WORK UNIT NO.:

11. TITLE (Include Security Classification): "Estimating Weights in Heteroscedastic Regression Models by Applying Least Squares to Squares or Absolute Residuals."

12. PERSONAL AUTHOR(S): Carroll, Raymond J., Davidian, Marie and Ruppert, David

13a. TYPE OF REPORT	13b. TIME COVERED	14. DATE OF REPORT (Yr., Mo., Day)	15. PAGE COUNT
technical	FROM 8/85 TO 8/86	October 1985	13

16. SUPPLEMENTARY NOTATION

17. COSATI CODES			18. SUBJECT TERMS (Continue on reverse if necessary and identify by block number)
FIELD	GROUP	SUB. GR.	

19. ABSTRACT (Continue on reverse if necessary and identify by block number)

We consider a nonlinear regression model for which the variances depend on a parametric function of known variables. We focus on estimating the variance function, after which it is typical to estimate the mean function by weighted least squares. Most often, squared residuals from an unweighted least squares fit are compared to their expectations and used to estimate the variance function. If properly weighted such methods are asymptotically equivalent to normal-theory maximum likelihood. Instead, one could use the deviations of the absolute residuals from their expectations. We construct such an estimator of the variance function based on absolute residuals whose asymptotic efficiency relative to maximum likelihood is precisely the same for symmetric errors as the asymptotic efficiency in the one-sample problem of the mean absolute deviation relative to the sample variance. The estimators are computable using nonlinear least squares software. The results hold with minimal distributional assumptions.

20. DISTRIBUTION/AVAILABILITY OF ABSTRACT	21. ABSTRACT SECURITY CLASSIFICATION
UNCLASSIFIED/UNLIMITED ☒ SAME AS RPT. ☐ DTIC USERS ☐	Unclassified

22a. NAME OF RESPONSIBLE INDIVIDUAL	22b. TELEPHONE NUMBER (Include Area Code)	22c. OFFICE SYMBOL
[redacted]	[redacted]	

DD FORM 1473, 83 APR EDITION OF 1 JAN 73 IS OBSOLETE.

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE

ESTIMATING WEIGHTS IN HETEROGENEOUS REGRESSION MODELS BY APPLYING LEAST SQUARES TO SQUARED OR ABSOLUTE RESIDUALS

Raymond J. Carroll, Marie Davidian

and

David Ruppert

University of North Carolina
Chapel Hill, N.C. 27514

Mimeo Series #1584

October 1985

DEPARTMENT OF STATISTICS
Chapel Hill, North Carolina

ESTIMATING WEIGHTS IN HETEROSCEDASTIC REGRESSION MODELS BY APPLYING LEAST SQUARES TO SQUARED OR ABSOLUTE RESIDUALS

Raymond J. Carroll

Marie Davidian

and

David Ruppert

Statistics Department

University of North Carolina

at Chapel Hill

Chapel Hill, N.C. 27514

Abstract

We consider a nonlinear regression model for which the variances depend on a parametric function of known variables. We focus on estimating the variance function, after which it is typical to estimate the mean function by weighted least squares. Most often, squared residuals from an unweighted least squares fit are compared to their expectations and used to estimate the variance function. If properly weighted such methods are asymptotically equivalent to normal-theory maximum likelihood. Instead, one could use the deviations of the absolute residuals from their expectations. We construct such an estimator of the variance function based on absolute residuals whose asymptotic efficiency relative to maximum likelihood is precisely the same for symmetric errors as the asymptotic efficiency in the one-sample problem of the mean absolute deviation relative to the sample variance. The estimators are computable using nonlinear least squares software. The results hold with minimal distributional assumptions.

1. Introduction

Consider a possibly nonlinear regression model in which the variance of the responses are not constant, but are parametric functions of predictor variables. More specifically, we have independent observations $Y_1, Y_2, \ldots, Y_N$ following the mean-variance model

$$E\, Y_i = f(x_i, \beta)$$
$$\text{Variance}\,(Y_i) = \xi\, h(z_i, \theta)\ . \tag{1.1}$$

In (1.1), $\{x_i, z_i\}$ are fixed constants, β is the vector regression parameter, θ is a vector of q components, f is called the regression function and h is called the variance function. See Judge, et al. (1985, Chapter 11) for a recent theoretical discussion of this basic heteroscedastic regression model, and Montgomery & Peck (1982, pages 99-104) for a simple example. If the structural parameter θ were known, redefining $Y_{i*} = Y_i/h^{1/2}(z_i,\theta)$ and $f_*(z_i, x_i, \beta) = f(x_i,\beta)/h^{1/2}(z_i,\theta)$ would yield a homoscedastic regression model which can be fit by one's favorite method.

We are interested in the case that the structural parameter θ is unknown. Given an estimate $\hat{\theta}$ of θ, the usual device for estimating the regression parameter β is simply to pretend that θ is known and equal to $\hat{\theta}$, and then proceed as in the previous paragraph. The resulting estimate of β will be called generalized least squares. It is one of the great folklore theorems of statistics which assures us that for estimating β, it really

does not matter how we estimate θ, at least asymptotically. More precisely, for large sample sizes the limiting distribution of generalized least squares is the same as if θ were known.

Despite the folklore asymptotics, as intuition would indicate for finite samples how one estimates θ really matters. Williams (1975) states that "both analytic and empirical studies of a variety of linear models indicate that ... the ordering by efficiency of (estimator of β) ... in small samples is in accordance with the ordering by efficiency (of estimates of θ)". In the linear model, Toyooka (1982), Rothenberg (1984) and Kariya (1985) all essentially show that for normally distributed data, the second order covariance matrix of generalized least squares is a monotonically increasing function of the covariance of the estimate of θ; see also Freedman & Peters (1984) and Carroll & Ruppert (1985) for similar results. Finally, especially the Monte-Carlo study of Goldfeld & Quandt (1972, pages 96-120) shows that it is possible to construct a disastrously inefficient generalized least squares estimator as well as quite an efficient one.

The purpose of this paper is to compare various estimators of θ by asymptotic efficiency. Without making any further assumptions than the minimal (1.1), it is possible to construct consistent and asymptotically normal estimates of θ with the following "algorithm":

(1.2) (1) Estimate β, obtain $\hat{\beta}$;
(2) Form squared residuals $r_i^2 = (Y_i - f(x_i, \hat{\beta}))^2$;
(3) Estimate θ by a function of the squared residuals.

Three common methods have been proposed, see Hildreth & Houck (1976), Amemiya (1977), Dent & Hildreth (1977), Jobson & Fuller (1980), Goldfeld & Quandt (1972), Harvey (1976) and Theil (1971), among others. The first

method is based on pretending that the data are normally distributed, in which case we can compute the maximum likelihood estimator $\hat{\theta}_{ML}$. If $\hat{\beta}$ in (1.2) is the maximum likelihood estimator, then $\hat{\theta}_{ML}$ solves

$$(1.3) \qquad 0 = N^{-1/2} \sum_{i=1}^{N} \left\{ \frac{r_i^2}{\hat{\xi}\, h(z_i,\hat{\theta})} - 1 \right\} \begin{pmatrix} 1 \\ \hat{\xi}\, v_i(\hat{\theta}) \end{pmatrix}$$

where

$$v_i = v_i(\theta) = \frac{\partial}{\partial \theta} \log h(z_i,\theta) .$$

Actually, the asymptotic distribution of solutions to (1.3) remains the same if the estimator of β satisfies

$$(1.4) \qquad N^{1/2}(\hat{\beta}-\beta) = O_p(1) ,$$

a fact sketched in the appendix. The logic of (1.3) is that $Er_i^2 \doteq \xi h(z_i,\theta)$. The other methods are also based on the idea that $\xi h(z_i,\theta)$ is approximately the expectation of squared residuals, see Jobson & Fuller (1980). For simplicity of presentation, we will ignore the asymptotically negligible bias in the squared residuals due to leverage. The unweighted estimate $\hat{\theta}_{LS}$ minimizes on (ξ,θ)

$$(1.5) \qquad \sum_{i=1}^{N} \{r_i^2 - \xi h(z_i,\theta)\}^2 ,$$

while the weighted estimate $\hat{\theta}_{WLS}$ minimizes in (ξ,θ)

$$(1.6) \qquad \sum_{i=1}^{N} \{r_i^2 - \xi h(z_i,\theta)\}^2 / h^2(z_i,\theta_{LS}) .$$

This last estimator is motivated by the idea that the variance of squared residuals is approximately proportional to $h^2(z_i,\theta)$, so that some sort of weighting ought to be employed. Also note that differentiating (1.6) yields

$$0 = \sum_{i=1}^{N} \left\{ \frac{r_i^2}{\hat{\xi} h(z_i,\hat{\theta})} - 1 \right\} \begin{pmatrix} 1 \\ \hat{\xi} v_i(\hat{\theta}) \end{pmatrix} \frac{h^2(z_i,\hat{\theta})}{h^2(z_i,\hat{\theta}_{LS})} ,$$

which differs from the likelihood equation (1.3) only by the asymptotically negligible factors $[h(z_i,\hat{\theta})/h(z_i,\hat{\theta}_{LS})]^2$. In Theorem 1 of the next section, we investigate these three estimators of θ, proving that (1) for all three the estimate of β is immaterial asymptotically as long as (1.4) holds; (2) maximum likelihood $\hat{\theta}_{ML}$ and weighted squared residuals $\hat{\theta}_{WLS}$ have the same limit distributions; and (3) both $\hat{\theta}_{ML}$ and $\hat{\theta}_{WLS}$ are asymptotically more efficient than using unweighted squared residuals via $\hat{\theta}_{LS}$. These results are obtained essentially independently of the underlying distribution.

Squared residuals are skewed and long-tailed. For this reason, Cohen, et al. (1984) suggest the use of absolute residuals, although they use absolute residuals only as one part of their algorithm and eventually use squared residuals. For the special case that Variance $(Y_i) = g(Z_i,\theta) = (Z_i^T\theta)^2$, apparently for computational reasons Glejser (1969) and Theil (1971) also propose use of absolute residuals. Such use requires a further assumption, namely that

$$(1.7) \qquad E|Y_i - f(x_i,\beta)| = \eta\, h^{1/2}(Z_i,\theta) .$$

Effectively, (1.1) and (1.7) require that

$$(1.8) \qquad \varepsilon_i = \frac{Y_i - f(x_i,\beta)}{\{\xi\, h(z_i,\theta)\}^{1/2}}$$

be independent and identically distributed, an assumption we shall make from now on.

Mimicking (1.5) and (1.6), one can construct two estimators of θ based on absolute residuals. Noting from (1.7)-(1.8) that absolute residuals have approximate expectation $\eta\, h^{1/2}(z_i,\theta)$ and variance proportional to $h(z_i,\theta)$, the unweighted absolute residual estimator $\hat{\theta}_{AV}$ minimizes in (η,θ)

$$(1.9) \qquad \sum_{i=1}^{N} \{|r_i| - \eta\, h^{1/2}(z_i,\theta)\}^2$$

while the weighted version $\hat{\theta}_{WAV}$ minimizes

$$(1.10) \qquad \sum_{i=1}^{N} \{|r_i| - \gamma h^{1/2}(z_i,\theta)\}^2/h(z_i,\hat{\theta}_{AV}) .$$

For the special case that the standard deviation is linear in exogenous variables, Judge, et al. (1985) propose our general absolute residual estimators. Even in this special case, they state that the properties of $\hat{\theta}_{AV}$ and $\hat{\theta}_{WAV}$ "have not been fully investigated". In their specific context, they go on to make in effect three conjectures:

(a) Absolute residual estimators of θ are not affected by the method of estimating β, as long as (1.4) holds;

(b) Weighted absolute residuals $\hat{\theta}_{WAV}$ are more efficient than not weighting and using $\hat{\theta}_{AV}$;

(c) If we define

$$\delta = \mathrm{Var}(|\varepsilon|)$$

$$\kappa = \mathrm{Var}(\varepsilon^2) ,$$

then in the light of Theorem 1 the asymptotic relative efficiency of the weighted absolute residual estimator $\hat{\theta}_{WAV}$ with respect to maximum likelihood $\hat{\theta}_{ML}$ or weighted squared residuals $\hat{\theta}_{WLS}$ is

$$(1.11) \qquad \frac{\kappa(1-\delta)}{4\delta} .$$

In this paper, we verify all these conjectures when the errors (1.8) are symmetrically distributed. In Section 3, we discuss why it is that, for this special case, using absolute residuals may be preferred when viewed from a perspective of efficiency robustness.

We also show that conjecture (a) and hence conjecture (c) are false in general for asymmetrically distributed errors. While the dependence of the asymptotic distribution on the estimate of β certainly complicates the theory, the dependence does not disqualify using absolute residuals. We exhibit a simple example for which using absolute residuals is always more than twice as efficient as using squared residuals.

The theorems are stated in the next section, with proofs in the appendix. In the third section we discuss the statistical implications of the results.

2. Major Results

We state the results somewhat informally and in the appendix only sketch the proofs, relying for the most part on simple Taylor series expansions and the somewhat more complex linearizations in Ruppert & Carroll (1980) and Carroll & Ruppert (1982). We first consider the estimators $\hat{\theta}_{ML}$, $\hat{\theta}_{LS}$ and $\hat{\theta}_{WLS}$ based on squared residuals. Under minimal assumption such as (1.1) and (1.4), these estimators are consistent and asymptotically normally distributed with asymptotic covariance unaffected by the choice of $\hat{\beta}$. The covariance simplifies if the errors (1.8) are independent and identically distributed, an assumption we will make throughout. Define

$$\Sigma_{ML} = \{N^{-1} \sum_{i=1}^{N} (v_i - \bar{v})(v_i - \bar{v})^T\}^{-1} .$$

Theorem #1. Under (1.4) and further regularity conditions, maximum likelihood $\hat{\theta}_{ML}$ and weighted squared residuals $\hat{\theta}_{WLS}$ have the same asymptotic distributions, with

$$N^{1/2}(\hat{\theta}_{ML} - \theta) \xRightarrow{\mathcal{L}} N(0, \Sigma_{ML})$$

$$N^{1/2}(\hat{\theta}_{WLS} - \theta) \xRightarrow{\mathcal{L}} N(0, \Sigma_{ML}) ,$$

where "$\xRightarrow{\mathcal{L}}$" means convergence in distribution. Further, unweighted squared residuals satisfies

$$N^{1/2}(\hat{\theta}_{LS} - \theta) \xRightarrow{\mathcal{L}} N(0, \Sigma_{LS}) ,$$

where $\Sigma_{LS} \geq \Sigma_{ML}$.

Thus, Theorem #1 assures us that we should weight the squared residuals for greatest asymptotic efficiency, at least when the errors are independent and identically distributed. The next result relates to the case of symmetric errors (1.8), thereby proving the three conjectures of Judge, et al. (1985) in this special case.

Theorem #2. Suppose that the errors (1.8) are symmetrically distributed with a distribution function which is continuous in a neighborhood of zero. Under (1.4) and further regularity conditions,

$$N^{1/2}(\hat{\theta}_{AV}-\theta) \Rightarrow N(0,\Sigma_{AV})$$

and

$$N^{1/2}(\hat{\theta}_{WAV}-\theta) \Rightarrow N(0,\Sigma_{WAV}) .$$

Further, $\Sigma_{WAV} \leq \Sigma_{AV}$, and

$$\Sigma_{WAV} = \frac{1}{(1-\delta)} \Sigma_{ML} .$$

If the errors are not symmetrically distributed, absolute residual estimation of θ is affected by how one estimates β, see Carroll & Schneider (1985) for a similar example of this phenomenon. Define

$$\gamma = \Pr(\varepsilon \geq 0) - \Pr(\varepsilon < 0) .$$

Recall that, by assumption, θ is a vector of q components. Define the matrices

$$C_{p^*} = N^{-1} \sum_{i=1}^{N} \begin{Bmatrix} 1 \\ \gamma v_i/2 \end{Bmatrix} (1, \gamma v_i^T/2) h^{p-1}(z_i,\theta)$$

$$\dot{f}(x_i,\beta) = \frac{\partial}{\partial\beta} f(x_i,\beta)$$

$$e = (0(q \times 1)\ I_q),$$

where I_q is the $q \times q$ identity matrix. Finally, define

$$\ell_i = \begin{pmatrix} 1 \\ \eta\, v_i/2 \end{pmatrix}$$

Theorem #3. Under (1.4) and further regularity conditions, we have the asymptotic expansions

$$N^{1/2}(\hat{\theta}_{AV}-\theta)$$

$$\doteq e\, C_{2^*}^{-1} \begin{Bmatrix} \xi^{1/2}\, N^{-1/2} \sum_{i=1}^{N} h(z_i,\theta)(|\varepsilon_i| - E|\varepsilon|)\ell_i \\ -\gamma\, N^{-1} \sum_{i=1}^{N} \ell_i\, h^{1/2}(z_i,\theta)\dot{f}^T(x_i,\beta)N^{1/2}(\hat{\beta}-\beta) \end{Bmatrix}$$

and

$$N^{1/2}(\hat{\theta}_{WAV}-\theta)$$

$$\doteq e\, C_{1^*}^{-1} \begin{Bmatrix} \xi^{1/2}\, N^{-1/2} \sum_{i=1}^{N} \ell_i(|\varepsilon_i| - E|\varepsilon|) \\ -\gamma\, N^{-1} \sum_{i=1}^{N} \ell_i\, h^{-1/2}(z_i,\theta)\dot{f}^T(x_i,\beta)N^{1/2}(\hat{\beta}-\beta) \end{Bmatrix}$$

Theorem #2 is a corollary of Theorem #3 because, under symmetry, $\gamma = 0$ and the effect of $\hat{\beta}$ disappears. In general, when $\gamma \neq 0$ $\hat{\theta}_{AV}$ and $\hat{\theta}_{WAV}$ are still asymptotically normally distributed, but their covariance matrices will depend on the method used to estimate β.

3. Discussion

In Section 2, we have shown that for estimating the structural parameter θ in the variance function, normal theory maximum likelihood is asymptotically equivalent to weighting squared residuals and applying a nonlinear least squares algorithm. This result holds essentially independently of the underlying distributions of the errors $\{\varepsilon_i\}$ in (1.8), which need not even be identically distributed. If the errors are identically distributed, then both methods are asymptotically more efficient than ordinary least squares. In practice, this means that if computing maximum likelihood is inconvenient as in Froehlich (1973) or Dent & Hildreth (1977), then in fitting squared residuals one ought to weight.

We have also shown that, if the errors (1.8) are independent and identically distributed symmetric random variables, then by appropriate weighting one can construct an estimator $\hat{\theta}_{WAV}$ which has asymptotic efficiency (1.11) relative to maximum likelihood $\hat{\theta}_{ML}$ and weighted squared residuals $\hat{\theta}_{WLS}$. For symmetric distributions, in one sample problems (1.11) is the asymptotic relative efficiency of the mean absolute deviation with respect to the sample variance, see Huber (1981, pages 2-3). For normally distributed data, using absolute residuals is 12% less efficient than using squared residuals. However, for the longer-tailed double exponential distribution, using absolute residuals is 25% more efficient. Huber (1981, page 3) presents an interesting computation of (1.5) for the class of contaminated normal distributions

$$(1 - \alpha)\Phi(\varepsilon) + \alpha\,\Phi(\varepsilon/3)\,,$$

where Φ is the normal distribution function. This distribution arises when a random fraction α of clean normally distributed data is contaminated by normal data with three times larger standard deviation, and it is commonly used in robustness studies. The relative efficiency of absolute values as a function of the contamination fraction α is given as follows:

α	Relative efficiency
0.0	87.6%
0.001	94.8%
0.002	101.6%
0.01	143.9%
0.05	203.5%

Huber calls these numbers "disquieting", noting that just 2 "bad" observations in 1000 suffice to offset the superiority of squared over absolute residuals when estimating the variance function.

If the errors are symmetrically distributed or nearly so, then robustness of efficiency considerations strongly suggest using weighted absolute residuals to estimate the variance function rather than weighted squared residuals or normal theory maximum likelihood. Computation is not intrinsically difficult since it is based on the usual nonlinear least squares methodology.

The residuals are defined through an estimate of the regression parameter β. The estimation of the variance function using squared residuals is asymptotically unaffected by the estimate of β. The same can be said for absolute residuals only when the errors are symmetrically distributed. Clearly, the use of absolute residuals is complicated and more research is needed in this direction. That further research may be quite useful is seen in the following

example. Let $\{W_i\}$ be independent and identically distributed negative exponential random variables with mean one. Consider the model of heteroscedastic regression through the origin,

$$Y_i = x_i \beta + \{\xi x_i^{\theta}\}^{1/2}(W_i-1) .$$

In this case, $\kappa = 9$, $v_i = \log x_i$ and writing

$$\mathrm{Var}(v) = N^{-1} \sum_{i=1}^{N} (v_i-\bar{v})^2 ,$$

normal theory maximum likelihood satisfies

$$N^{1/2}(\hat{\theta}_{ML}-\theta) \overset{\mathcal{L}}{\Rightarrow} N\big(0,9.0/\mathrm{Var}(v)\big).$$

When $\theta = 2$, simple calculations show that the estimate of β does not matter and that

$$N^{1/2}(\hat{\theta}_{WAV}-\theta) \overset{\mathcal{L}}{\Rightarrow} N\big(0,3.4/\mathrm{Var}(v)\big) .$$

Writing

$$a_0 = N^{-1} \sum_{i=1}^{N} x_i^{2-\theta}$$

$$b_0 = N^{-1} \sum_{i=1}^{N} (v_i-v)x_i^{1-\theta/2} ,$$

we find that if $\hat{\beta}$ is any generalized least squares estimate, then from the appendix,

$$N^{1/2}(\hat{\theta}_{WAV}-\theta) \Rightarrow N\left(0, \frac{3.4}{\mathrm{Var}(v)} + \frac{2.37\ b_0^2}{(\mathrm{Var}(v))^2\ a_0}\right)$$

indicating the asymptotic superiority of using absolute residuals as long as

(3.1) $\qquad b_0^2 \leq 2.36\, a_0\, \mathrm{Var}(v)$

By the Cauchy-Schwarz inequality, $b_0^2 \leq a_0\, \mathrm{Var}(v)$ and from (3.1) we see that using absolute residuals will always be more than two times as efficient as the MLE or squared residuals.

The point of the previous example is that absolute residuals estimation of θ should not be automatically dismissed simply because it has an inconvenient asymptotic theory under asymmetric errors. As long as one can reasonably make the crucial assumption (1.7), using weighted absolute residuals to estimate the variance function should be given serious consideration. However, further research is needed to help the statistician choose between using weighted squared or absolute residuals when asymmetry is present.

We have confined our discussion to weighted least squares estimation of β and absolute versus squared residuals for estimating the variance function. Our techniques apply to other methods, including using weighted logarithms of squared residuals and the robust estimation schemes of Carroll & Ruppert (1982) and Giltinan, Carroll & Ruppert (1986).

ACKNOWLEDGEMENT

The work of Carroll and Davidian was supported by the Air Force Office of Scientific Research Contract AFOSR-F49620-82-C-0009. Ruppert was supported by National Science Foundation Grant MCS-8100748.

REFERENCES

Amemiya, T. (1977). A note on a heteroscedastic model. Journal of Econometrics 6, 365-370 and corrigenda 8, 275.

Carroll, R.J. & Ruppert, D. (1982). Robust estimation in heteroscedastic linear models. Annals of Statistics 10, 429-441.

Carroll, R.J. & Ruppert, D. (1985). A note on the effect of estimating weights in weighted least squares. Preprint.

Carroll, R.J. & Schneider, H. (1985). A note on Levene's tests for heteroscedasticity. Letters in Probability and Statistics, to appear.

Cohen, M.L., Dalal, S.R. & Tukey, J.W. (1984). Robust, smoothly heterogeneous variance regression. Preprint.

Dent, W.T. & Hildreth, C. (1977). Maximum likelihood estimation in random coefficient models. Journal of the American Statistical Association 72, 69-72.

Freedman, D.A. & Peters, S.C. (1984). Bootstrapping a regression equation: some empirical results. Journal of the American Statistical Association 79, 97-104.

Froehlich, B.R. (1973). Some estimators for a random coefficient regression model. J. Am. Statist. Assoc. 68, 329-335.

Giltinan, D. M., Carroll, R. J. & Ruppert, D. (1986). Some new heteroscedastic regression methods when there are possible outliers. Technometrics, to appear.

Glejser, H. (1969). A new test for heteroscedasticity. Journal of the American Statistical Association 64, 316-323.

Goldfeld, S.M. & Quandt, R.E. (1972). Nonlinear Methods in Econometrics. North Holland, Amsterdam.

Harvey, A.C. (1976). Estimating regression models with multiplicative heteroscedasticity. Econometrics 44, 461-465.

Hildreth, C. & Houck, J.P. (1968). Some estimators for a linear model with random coefficients. Journal of the American Statistical Association 63, 584-595.

Huber, P.J. (1981). Robust Statistics. John Wiley & Sons, New York.

Jobson, J.D. & Fuller, W.A. (1980). Least squares estimation when the covariance matrix and parameter vector are functionally related. Journal of the American Statistical Association 75, 176-181.

Judge, G.G., Griffiths, W.E., Hill, R.C., Lutkepohl, H. & Lee, T.C. (1985). The Theory and Practice of Econometrics, Second Edition. John Wiley & Sons, New York.

Kariya, T. (1985). A nonlinear version of the Gauss-Markov theorem. Journal of the American Statistical Association 80, 476-477.

Montgomery, D.C. & Peck, E.A. (1982). Introduction to Linear Regression Analysis. John Wiley & Sons, New York.

Rothenberg, T.J. (1984). Approximate normality of generalized least squares estimates. Econometrika 52, 811-825.

Ruppert, D. & Carroll, R.J. (1980). Trimmed least squares estimation in the linear model. Journal of the American Statistical Association 75, 828-838.

Theil, H. (1971). Principles of Econometrics. John Wiley & Sons, New York.

Toyooka, Y. (1982). Second order expansion of mean squared error matrix of generalized least squares estimator with estimated parameters. Biometrika 69, 269-273.

Williams, J. S. (1975). Lower bounds on convergence rates of weighted least squares to best linear unbiased estimators. In A Survey of Statistical Design and Linear Models, J.N. Srivastava, editor. Amsterdam: North Holland.

Appendix

Proofs of the main results

To keep this section somewhat self-contained, there is some redundancy with the text. Let $\hat{\beta}$ be any estimator satisfying

(A.1) $$N^{1/2}(\hat{\beta} - \beta) = O_p(1) .$$

Let $\{a_i\}$ be any sequence of constants. Define

$$r_i = Y_i - f(x_i, \hat{\beta})$$

$$s_i = Y_i - f(x_i, \beta) = \{\sigma^2 h(z_i, \theta)\}^{1/2}\varepsilon_i$$

$$\gamma = Pr(\varepsilon > 0) - Pr(\varepsilon < 0)$$

$$c_i = \frac{\partial}{\partial \beta} f(x_i, \beta) = \dot{f}(x_i, \beta)$$

$$b_i = a_i\{\sigma^2 h(z_i, \theta)\}^{1/2}$$

$$d_i = c_i\{\sigma^2 h(z_i, \theta)\}^{-1/2} .$$

Lemma #1. Under regularity conditions,

(A.2) $$N^{-1/2} \sum_{i=1}^{N} a_i r_i^2 = N^{-1/2} \sum_{i=1}^{N} a_i s_i^2 + o_p(1)$$

Proof:

This follows because

$$N^{-1/2} \sum_{i=1}^{N} a_i(r_i^2 - s_i^2)$$

$$= N^{-1/2} \sum_{i=1}^{N} a_i\{f(x_i, \hat{\beta}) - f(x_i, \beta)\}^2$$

$$-2 N^{-1/2} \sum_{i=1}^{N} a_i\{[\sigma^2 h(z_i, \theta)]^{1/2} [f(x_i, \hat{\beta}) - f(x_i, \beta)]\}$$

$$\xrightarrow{p} 0$$

by Taylor series, (A.1) and the fact that $E\varepsilon = 0$. From (A.2), we see that in computing estimates of θ based on squared residuals, it is sufficient to do the asymptotic distribution theory assuming β is known.

Proposition #1. If the distribution function $\{\varepsilon_i\}$ is continuous at zero, its mean, then

$$\lim_{N\to\infty} N^{1/2} E\{|\varepsilon - v/N^{1/2}| - |\varepsilon|\} = -v\nu .$$

Proof: Routine.

Lemma #2. Make the assumption of Proposition #1. Define

$$H_N = N^{-1/2} \sum_{i=1}^{N} a_i |s_i|$$

$$- \gamma N^{-1} \sum_{i=1}^{N} a_i \dot{f}(x_i, \beta)^T N^{1/2}(\hat{\beta}-\beta) .$$

Then

$$(A.3) \qquad H_N - N^{-1/2} \sum_{i=1}^{N} a_i |r_i| = o_p(1) .$$

Proof: Define

$$(A.4) \qquad Q_N(\Delta) = N^{-1/2} \sum_{i=1}^{N} a_i \left\{ |s_i - f(x_i, \beta + \Delta/N^{1/2}) + f(x_i, \beta)| - |s_i| \right\}$$

As in Ruppert & Carroll (1980) or Carroll & Ruppert (1982), for every $M > 0$ we have

$$\sup_{\|\Delta\| \le M} |Q_N(\Delta) - E\, Q_N(\Delta)| = o_p(1) .$$

Writing

$$N^{1/2}(f(x_i, \beta + \Delta/N^{1/2}) - f(x_i, \beta))$$

$$\approx c_i^T \Delta, \text{ where } c_i = f_\beta(x_i, \beta) ,$$

we see from Proposition #1 that

$$\text{(A.5)} \qquad E\, Q_N(\Delta) \quad -\gamma N^{-1} \sum_{i=1}^{N} a_i \dot{f}(x_i, \beta)^T \Delta .$$

Substituting $N^{1/2}(\hat{\beta}-\beta)$ for Δ and putting (A.5) into (A.4) shows that

$$\text{(A.6)} \qquad N^{-1/2} \sum_{i=1}^{N} a_i |r_i| + \gamma N^{-1} \sum_{i=1}^{N} a_i \dot{f}(x_i, \beta)^T N^{1/2}(\hat{\beta}-\beta)$$

$$-N^{-1/2} \sum_{i=1}^{N} a_i |s_i| = o_p(1) ,$$

completing the proof. Lemma #2 will assure us that, when dealing with methods based on absolute residuals, we may replace

$$N^{-1/2} \sum_{i=1}^{N} a_i |r_i|$$

by

$$N^{-1/2} \sum_{i=1}^{N} a_i |s_i| - \gamma N^{-1} \sum_{i=1}^{N} a_i \dot{f}(x_i, \beta)^T N^{1/2}(\hat{\beta}-\beta) .$$

This makes the proofs routine, and eliminates the effect of the estimate of β when $\gamma = 0$, see Theorem #2. Define

$$v_i = v_i(\theta) = h(Z_i, \theta)/h(Z_i, \cdot)$$

$$C_p = N^{-1} \sum_{i=1}^{N} \begin{Bmatrix} 1 \\ v_i \end{Bmatrix} (1 \; v_i^T)\, h^{2(p-1)}(Z_i, \theta)$$

$$C_{p*} = N^{-1} \sum_{i=1}^{N} \begin{Bmatrix} 1 \\ r\, v_i/2 \end{Bmatrix} (1 \; r v_i^T/2)\, h^{p-1}(Z_i, \theta)$$

Proof of Theorem #1. We will study each estimator in turn, only sketching the proof. For typing convenience, we will use the generic (ξ, θ), which will refer in turn to the estimator under consideration. Because of Lemma #1, we may assume that β is known.

Maximum Likelihood. Using a Taylor series and Lemma #1, we have that

$$o = N^{-1/2} \sum_{i=1}^{N} \left\{ \frac{r_i^2}{\hat{\xi}\, h(Z_i, \hat{\theta})} - 1 \right\} \begin{pmatrix} 1 \\ \hat{\xi}\, v_i(\hat{\theta}) \end{pmatrix}$$

$$\doteq N^{-1/2} \sum_{i=1}^{N} \left\{ \frac{r_i^2}{\xi\, h(Z_i, \theta)} - 1 \right\} \begin{pmatrix} 1 \\ \xi\, v_i \end{pmatrix}$$

$$- N^{-1} \sum_{i=1}^{N} \begin{pmatrix} 1 \\ \xi\, v_i \end{pmatrix} (1/\xi\, v_i^T)\, N^{1/2} \begin{pmatrix} \hat{\xi} - \xi \\ \hat{\theta} - \theta \end{pmatrix}$$

$$\doteq N^{-1/2} \sum_{i=1}^{N} (\varepsilon_i^2 - 1) \begin{pmatrix} 1 \\ \xi\, v_i \end{pmatrix} - \frac{1}{\xi}\, C_1\, N^{1/2} \begin{pmatrix} \hat{\xi} - \xi \\ \hat{\theta} - \theta \end{pmatrix}$$

Thus,

$$N^{1/2} \begin{pmatrix} \hat{\xi} - \xi \\ \hat{\theta} - \theta \end{pmatrix} \Longrightarrow N(0,\ \xi^2\, K\, C_1^{-1})$$

Easy algebra yields the result.

Unweighted squared Residuals. Again, from Lemma #1,

$$o = N^{-1/2} \sum_{i=1}^{N} h(Z_i, \hat{\theta}) \{r_i^2 - \hat{\xi}\, h(Z_i, \hat{\theta})\} \begin{pmatrix} 1 \\ \hat{\xi}\, v_i(\hat{\theta}) \end{pmatrix}$$

$$\doteq N^{-1/2} \sum_{i=1}^{N} \xi\, h^2(Z_i, \theta)\, (\varepsilon_i^2 - 1) \begin{pmatrix} 1 \\ \xi\, v_i \end{pmatrix}$$

$$- N^{-1} \sum_{i=1}^{N} h^2(Z_i, \theta) \begin{pmatrix} 1 \\ \xi\, v_i \end{pmatrix} (1 \quad \xi\, v_i^T)\, N^{1/2} \begin{pmatrix} \hat{\xi} & \xi \\ \hat{\theta} & \theta \end{pmatrix}$$

so that

$$N^{1/2} \begin{pmatrix} \hat{\xi} - \xi \\ \hat{\theta} - \theta \end{pmatrix} \xrightarrow{\mathcal{L}} N(0,\ \xi^2\, K\, C_2^{-1} C_3 C_2^{-1})$$

Since $C_1^{-1} \leq C_2^{-1}\, C_3\, C_2^{-1}$, unweighted squared residuals are less efficient than maximum likelihood.

Weighted Squared Residuals. Again, from Lemma #1,

$$o = N^{-1/2} \sum_{i=1}^{N} \begin{pmatrix} \dfrac{h(Z_i, \hat{\theta})}{h(Z_i, \hat{\theta}_{LS})} \end{pmatrix} \begin{pmatrix} \dfrac{r_i^2 - \xi\, h(Z_i, \hat{\theta})}{h(Z_i, \hat{\theta}_{LS})} \end{pmatrix} \begin{pmatrix} 1 \\ \xi\, v_i(\hat{\theta}) \end{pmatrix}$$

$$\doteq N^{-1/2} \sum_{i=1}^{N} \xi(\varepsilon_i^2 - 1) \begin{pmatrix} 1 \\ \xi\, v_i \end{pmatrix}$$

$$- C_1 N^{1/2} \begin{pmatrix} \hat{\xi} - \xi \\ \hat{\theta} - \theta \end{pmatrix}.$$

This shows, as claimed, that

$$N^{1/2} \begin{pmatrix} \hat{\xi} - \xi \\ \hat{\theta} - \theta \end{pmatrix} \Longrightarrow N(0,\ \xi^2\, K\, C_1^{-1}).$$

Proof of Theorem #3.

By Lemma #2, we will be able to replace $|r_i|$ by $|s_i| - \gamma\, \dot{f}(x_i, \beta)^T(\hat{\beta} - \beta)$. Recall that we are writing $E|\varepsilon| = \eta/\xi^{1/2}$, and that $[E|\varepsilon|]^2 = 1-\delta$, $\delta = \mathrm{Var}(|\varepsilon|)$. For unweighted absolute residuals, from Lemma #2 and a Taylor series we obtain

$$o \doteq N^{-1/2} \sum_{i=1}^{N} h^{1/2}(Z_i, \hat{\theta}) \left(|r_i| - \hat{\eta}\, h^{1/2}(Z_i, \hat{\theta})\right) \begin{pmatrix} 1 \\ \hat{\eta}\, v_i(\hat{\theta})/2 \end{pmatrix}$$

$$= \xi^{1/2} N^{-1/2} \sum_{i=1}^{N} \lambda_i\, h(Z_i, \theta)(|\varepsilon_i| - E|\varepsilon|)$$

$$- \gamma N^{-1} \sum_{i=1}^{N} \lambda_i\, h^{1/2}(Z_i, \theta)\, \dot{f}^T(x_i, \beta)\, N^{1/2}(\hat{\beta} - \beta)$$

$$- C_{2*} N^{1/2} \begin{pmatrix} \hat{\xi} - \eta \\ \hat{\theta} - \theta \end{pmatrix}$$

This is the first part of Theorem #3. Noting that for weighted absolute residuals

$$o = N^{-1/2} \sum_{i=1}^{N} \begin{pmatrix} \dfrac{h(z_i, \hat{\theta})}{h(Z_i, \hat{\theta}_{LS})} \end{pmatrix}^{1/2} \begin{pmatrix} \dfrac{|r_i| - \hat{\eta}\, h^{1/2}(Z_i, \hat{\theta})}{h^{1/2}(Z_i, \hat{\theta}_{LAV})} \end{pmatrix} \begin{pmatrix} 1 \\ \hat{\eta}\, v_i(\hat{\theta})/2 \end{pmatrix},$$

essentially the same application of Lemma #2 and Taylor series completes the proof.

Proof of Theorem #2. For the symmetric case, $\gamma = 0$. From Theorem #3,

$$N^{1/2}(\hat{\theta}_{AV} - \theta) \Longrightarrow N(0, \xi\ \delta\ C_{2*}^{-1}\ C_{3*}C_{2*}^{-1})$$

$$N^{1/2}(\hat{\theta}_{WAV} - \theta) \Longrightarrow N(0, \xi\ \delta\ C_{1*}^{-1}) .$$

Noting that $\xi = \eta^2/(1-\delta)$ and simple algebra completes the proof.

Proof of 3.2. Detailed calculations yield

$$\gamma = Pr(\varepsilon > 0) - Pr(\varepsilon < 0) = 2\ e^{-1} - 1$$

$$\delta = Var(|\varepsilon|) = 1 - 4\ e^{-2}$$

$$E|\varepsilon| = 2\ e^{-1}$$

$$E(\varepsilon|\varepsilon|) = 4\ e^{-1} - 1$$

$$e\ C_{1*}^{-1}\ \ell_i = \frac{2}{\eta\ Var(v)}\ (v_i - \bar{v}) .$$

For any generalized least squares estimate of β,

$$N^{1/2}(\hat{\beta} - \beta)/\xi^{1/2} \doteq (1/a_0)\ N^{-1/2} \sum_{i=1}^{N} x_i^{1-\theta/2}\ \varepsilon_i .$$

Substituting into Theorem #3 yields

$$N^{1/2}(\hat{\theta}_{WAV} - \theta)$$

$$\doteq \frac{2}{E|\varepsilon|\ Var(v)}\ N^{-1/2} \sum_{i=1}^{N} \begin{Bmatrix} (v_i - v)(|\varepsilon_i| - E|\varepsilon|) \\ -\gamma(b_0/a_0)\ x_i^{1-\theta/2}\ \varepsilon_i \end{Bmatrix}$$

The result now follows immediately.

A NOTE ON THE EFFECT OF IGNORING SMALL MEASUREMENT ERRORS IN PRECISION INSTRUMENT CALIBRATION

Raymond J. Carroll

and

Clifford H. Spiegelman

Raymond J. Carroll is Professor of Statistics at the University of North Carolina, Chapel Hill. Clifford H. Spiegelman is a Mathematical Statistician, Statistical Engineering Division, National Bureau of Standards.

Acknowledgement

This research was supported by the Air Force Office of Scientific Research Contract AFOSR F49620-82-C-0009 and by the Office of Naval Research Contracts N00014-83-K-0005 and NR-042544. The authors thank Dr. Robert Watters and Charles P. Reeve for helpful conversations, and the referee for a thorough review.

Abstract

Our focus is the simple linear regression model with measurement errors in both variables. It is often stated that if the measurement error in x is "small", then we can ignore this error and fit the model to data using ordinary least squares. There is some ambiguity in the statistical literature concerning the exact meaning of a "small" error. For example, Draper and Smith (1981) state that if the measurement error variance in x is small relative to the variability of the true x's, then "errors in the x's can be effectively ignored", see Montgomery & Peck (1983) for a similar statement. Scheffe (1973) and Mandel (1984) argue for a second criterion, which may be informally summarized that the error in x should be small relative to (the standard deviation of the observed Y about the line)/(slope of the line). We argue that for calibration experiments both criteria are useful and important, the former for estimation of x given Y and the latter for the lengths of confidence intervals for x given Y.

1. Introduction

There is substantial literature on the problem of precision instrument calibration, see for example Scheffe (1973), Rosenblatt and Spiegelman (1981) and Mandel (1984). We will focus on such calibration when fitting a straight line to a set of data in which the predictor x is measured with error.

Recently we were asked to try to quantify what is meant by a "small" measurement error in x, with the idea that, if such error were small, we could safely ignore it and procede with ordinary least squares analysis. In trying to do this we realized that the literature is somewhat ambiguous, and in fact there are two distinct criteria used to decide when measurement error in x is small. For example, Draper and Smith (1981, page 124) state that if the measurement error variance in x is small relative to the variability of the true x's themselves, then "errors in the x's can be effectively ignored and the usual least squares analysis performed". This comment is echoed by Montgomery and Peck (1982, page 388). On the other hand, both Scheffe (1973, page 2) and Mandel (1984) use the criterion that we can safely ignore measurement error in x if its standard deviation is small relative to the ratio

$$\frac{\text{Standard deviation of measured Y about the line.}}{\text{Slope of the line}}$$

The authors were working in different contexts, so it is not surprising that their criteria differ.

In this paper, we point out that for calibration experiments both criteria are useful. The criterion used by Draper and Smith is appropriate when the goal is estimation of intercept and slope based on the calibration data set, and then at the second stage for estimating the true value of x from a new observed Y. The criterion of Scheffe and Mandel addresses the issue of lengths of confidence intervals for estimating x from an observed Y. If the Draper and Smith criterion is satisfied while that of Scheffe and Mandel is not, the effect of ignoring the measurement error in x is essentially to cause larger confidence intervals for estimating the true value of x from new observed Y than is necessary.

Suppose that observed responses $\{Y_i\}$ are related linearly to the true working standards $\{x_i\}$ through the equation

$$Y_i = \alpha + \beta x_i + \epsilon_i \ , \quad i = 1,2,\ldots N. \tag{1.1}$$

Here the deviations $\{\epsilon_i\}$ combine measurement errors in the response with equation or model error, and the $\{\epsilon_i\}$ are normally distributed with mean zero and common variance σ_ϵ^2 .

Rather than observing the true working standards $\{x_i\}$, we observe

$$X_i = x_i + v_i \tag{1.2}$$

where the measurement errors $\{v_i\}$ are assumed normally distributed with mean zero and variance σ_m^2 . In the terminology of Fuller (1986), the equation (1.1) includes both equation error and response measurement error. From now on, when we speak of measurement error we will mean measurement error in the true $\{x_i\}$.

Assuming the working standards $\{x_i\}$ are measured without error, one would often procede as follows. First, perform the usual least squares analysis, which yields estimates $(\hat{\alpha}_L, \hat{\beta}_L, \hat{\sigma}_L)$. A new, independent observation Y_* is then made, and the goal is to estimate the value of x_* such that

$$E\, Y_* = \alpha + \beta\, x_* .$$

The maximum likelihood estimator is

$$\hat{x}_* = (Y_* - \hat{\alpha}_L)/\hat{\beta}_L. \tag{1.3}$$

For confidence intervals, the Working-Hotelling $100(1-\alpha)$ % interval (Seber (1977)) for the unknown x_* is

$$I = \{x:\ Y_* \text{ is contained in the interval } \hat{\alpha}_L + \hat{\beta}_L\, x \pm t_\alpha\, \hat{\sigma}_L\, R(x)\}, \tag{1.4}$$

where t_α is the $1-\alpha/2$ percentage point of the t-distribution with N-2 degrees of freedom, and

$$R^2(x) = 1 + N^{-1}\{1 + (x-\bar{x})^2/s_x^2\} ,$$

where $\bar{x}$, s_x^2 are given by

$$\bar{x} = N^{-1} \sum_{1}^{N} x_i \ , \quad s_x^2 = N^{-1} \sum_{1}^{N} (x - \bar{x})^2 .$$

If the calibration is to be repeated, more complex confidence statements are available for those who wish to use them, see Scheffé (1973).

Draper and Smith's criterion for the severity of measurement error is

$$\frac{\text{measurement error variance in the } \{x_i\}}{\text{Variation of the } \{x_i\}} \pm \frac{\sigma_m^2}{s_x^2} . \tag{1.5}$$

Scheffé and Mandel propose that the severity of measurement error depends on the size of

$$\sigma_m^2/(\sigma_e/\beta)^2 \ . \tag{1.6}$$

In the next section we discuss the criteria (1.5)-(1.6) with regard to estimation and confidence intervals for x_* given an observed Y_*.

2. The Effect of Small Error

The working standards $\{x_i\}$ are fixed constants, and the criterion (1.5) thus depends on the sample working standards. For large enough samples, we will think of the mean of the $\{x_i\}$ as converging to μ_x and the variance of the $\{x_i\}$ also converging, so that (1.5) can be written as

$$\lambda_N = \sigma_m^2/s_x^2 \simeq \lambda. \tag{2.1}$$

The least squares estimates $(\hat{\alpha}_L, \hat{\beta}_L)$ converge in probability to $(\alpha + \lambda \mu_x \beta/(1+\lambda), \beta/(1+\lambda))$ respectively. By centering appropriately so that $\mu_x \simeq 0$, we see that the bias in least squares essentially depends on the size of λ in (2.1). When λ is small, for the purpose of estimation, the effect of ignoring measurement error in the true $\{x_i\}$ is slight.

There is no standard method to correct for measurement error when estimating $(\alpha, \beta, \sigma_e, \sigma_m)$. For example, when there is no replication in the experiment, it is customary to assume that the ratio

$$\theta = \sigma_m^2/\sigma_e^2 \tag{2.2}$$

is known, see Kendall & Stuart (1961, pages 375-387) or Fuller (1986). In some applications, θ will be known from the physical set-up of the problem.

For the effect of misspecifying θ, see Lakshminarayanan & Gunst (1984) and Ketellapper (1983). The basic danger is in thinking that θ is larger than it actually is. In practice, if θ is not known one usually considers replicating the responses and/or the predictors so as to allow estimation of σ_m and σ_e, see Fuller (1986) for a thorough discussion.

Regardless of whether θ is known or replication is used, we can make the following general qualitative statement. When λ is small, not only are the least squares estimators nearly the same as the maximum likelihood estimators, but in particular the least squares estimators are approximately unbiased as discussed previously. The story is considerably different when we turn to confidence intervals. Define

L_1 = length of the confidence interval for x_* given Y_* taking into account the measurement error in $\{x_i\}$.

L_2 = length of the confidence interval for x_* ignoring the measurement error in the $\{x_i\}$.

If we assume that the sample sizes are large enough and, if replication is used, there are sufficient degrees of freedom in the replication, in Appendix A we verify that when λ is small the ratio of the confidence interval lengths is approximately

$$\frac{L_2}{L_1} \simeq \{1 + (\frac{\sigma_m}{\sigma_e/\beta})^2\}^{\frac{1}{2}}. \tag{2.3}$$

The reason that (2.3) holds is that, as seen in (1.4), the length L_2 of confidence interval ignoring measurement error is essentially proportional to $\hat{\sigma}_L$, which converges in probabilty to $(\sigma_e^2 + \beta^2\sigma_m^2)^{\frac{1}{2}}$, while the length L_1

is proportional to an estimate of σ_e; the ratio of these two lengths is (2.3).

Equation (2.3) verifies the criterion of Scheffe and Mandel that for confidence intervals, we can ignore measurement error in the working standards only if the measurement error has variance σ_m^2 small relative to σ_e^2/β^2. In the next section we provide an example where the criterion (1.5) mentioned by Draper & Smith is small but the Scheffe and Mandel criterion (1.6) is large.

3. An Example

In Table 1 we list a subset of the data investigated by Lechner, Reeve & Spiegelman (1982). It is not our purpose to provide a definitive analysis of these data. Rather, we use the data only to provide a means of exploring the effect of ignoring small measurement error, especially through the increased length ratio (2.3). We assume a straight line fit (1.1) to the data. We find that $\hat{\alpha}_L = -291.49$, $\hat{\beta}_L = 2346.64$ and $\hat{\sigma}_L = 1.64$. From discussion with the investigators it was thought that σ_m and σ_e are of the same order of magnitude. However, since σ_e is made up of both response measurement error and equation error, for this illustration we decided to be rather conservative as suggested by Lakshminarayanan & Gunst (1984) and Ketellapper (1983) and set $\theta = 0.001$ in (2.2). Following Kendall & Stuart (1961), the maximum likelihood estimators of (α,β,σ) assuming θ is known are given by

$$\hat{\alpha}_* = \bar{Y} - \hat{\beta}_* \bar{X}$$

$$\hat{\beta}_* = \frac{(S_Y^2 - \theta^{-1}S_X^2) + \{(S_Y^2 - \theta^{-1}S_X^2)^2 + 4\theta^{-1}S_{YX}^2\}^{\frac{1}{2}}}{2S_{YX}}$$

$$\hat{\sigma}_m = \theta^{-1}[S_X^2 - S_{YX}/\hat{\beta}_*],$$

where

$$S_X^2 = \frac{1}{N}\sum_{i=1}^{N}(x_i - \bar{x})^2$$

$$S_Y^2 = \frac{1}{N}\sum_{i=1}^{N}(Y_i - \bar{Y})^2$$

$$S_{YX} = \frac{1}{N}\sum_{i=1}^{N}(X_i - \bar{X})(Y_i - \bar{Y}).$$

It is known that the maximum likelihood estimator for σ is biased even in larger samples, and it is customary to make the correction

$$\hat{\sigma}_{m*} = 2\hat{\sigma}_m.$$

We found that

$$\hat{\beta}_* = 2346.64, \ \hat{\sigma}_{m*} = 6.77 \times 10^{-4}.$$

Making the rough approximations

$$\beta \simeq \hat{\beta}_*, \ \sigma_m \simeq \hat{\sigma}_{m*}, \ \sigma_m^2 \simeq .001\sigma_e^2 \quad \text{and}$$

$$\sigma_x^2 \simeq \text{Sample variance of observed X's} - \hat{\sigma}_m^2 \simeq 0.57,$$

we find that $\lambda \leq 0.001$. Since λ is very small and $\hat{\beta}_L \simeq \hat{\beta}_*$, we conclude that for purposes of estimation, measurement error in the $\{x_i\}$ can be

effectively ignored. However, the ratio of the lengths of the confidence intervals for x_* is approximately

$$L_2/L_1 \simeq (1 + \theta\, \beta^2)^{\frac{1}{2}} \simeq 74.2 \ .$$

This large ratio emphasizes our point that the definition of "small measurement error" must depend on whether one is interested in estimation or confidence intervals.

4. Conclusion

We have shown that, under the ideal conditions of a straight line model and a fairly large-sized working sample, ignoring measurement errors in x which are "small" relative to the usual estimation criterion (2.1) can result in calibration confidence intervals which are much larger than necessary. For confidence intervals, it is more sensible to judge measurement error size on the basis of both (1.5) and (2.3). Ignoring the measurement error in the true working standards $\{x_i\}$ will cause an increase in confidence interval length on the order of (2.3).

We finish by emphasizing that using measurement error techniques to obtain shorter calibration confidence intervals requires that equation (1.1) should hold. While least square confidence intervals can be very conservative in examples such as we have studied, they are more robust against small model misspecifications. Small perturbations from the straight-line fit can significantly alter the coverage probabilities of the measurement error confidence interval I_1 without greatly affecting the coverage of the least squares intervals.

References

Draper, N.R. & Smith, H. (1981). Applied Regression Analysis, second edition. John Wiley & sons, New York.

Fuller, W.A. (1986). Measurement Error Models. John Wiley & Sons, New York.

Kendall, M.G. & Stuart, A. (1961). The Advanced Theory of Statistics, Volume 2. Charles Griffin & Company.

Lechner, J.A., Reeve, C.P. & Spiegelman, C.H. (1982). An implementatin of the Scheffe approach to calibration using spline functions, illustrated by a pressure-volume calibration. Technometrics 24, 229-234.

Mandel, J. (1984). Fitting straight lines when both variables are subject to error. Journal of Quality Technology 16, 1-13.

Montgomery, D.C. & Peck, E.A. (1982). Introduction to Linear Regression Analysis. John Wiley & Sons, New York.

Rosenblatt, J.R. & Spiegelman, C.H. (1981). Discussion of "A Bayesian analysis of the linear calibration problem" by William G. Hunter and Warren F. Lamboy, Technometrics 23, 329-333.

Scheffe, H. (1973). A statistical theory of calibration. Annals of Statistics 1, 1-37.

Seber, G.A.F. (1977). Linear Regression Analysis. John Wiley & Sons, New York.

Appendix A

In this appendix, we verify the approximation (2.3). While a precise large-sample analysis is routine, it is also notationally quite cumbersome. The essential ideas are perhaps easier to understand through the following heuristic analysis. Suppose that N is large and that λ in (2.1) is small. Assuming that

(A.1) $$\sigma_m^2/\sigma_\epsilon^2 = \theta \quad \text{known},$$

then maximum likelihood estimates $(\hat{\alpha}_*, \hat{\beta}_*)$ can be formed which are consistent for (α,β), see Fuller (1986). Under the assumption of small λ and large sample size N, we have

$$\hat{\alpha}_L \simeq \hat{\alpha}_* \simeq \alpha; \quad \hat{\beta}_L \simeq \hat{\beta}_* \simeq \beta ;$$

$$R(x) \simeq 1; \quad \hat{\sigma}_{m*} \simeq \sigma_L, \ \hat{\sigma}_L \simeq (\sigma_\epsilon^2 + \beta^2 \sigma_m^2)^{\frac{1}{2}} .$$

Here $\hat{\sigma}_{m*}$ is the usual consistent estimate of σ_ϵ under the assumption (2.2). Taking into account the measurement error in $\{x_i\}$ and using $(\hat{\alpha}_*, \hat{\beta}_*, \hat{\sigma}_{m*})$, within our heuristic framework the appropriate Working-Hotelling confidence interval for x_* is approximately

$$I_1 = \{x: Y_* \in \hat{\alpha}_* + \hat{\beta}_* \pm z_\alpha \hat{\sigma}_{m*}\} ,$$

where z_α is the $1-\alpha/2$ standard normal percentage point. The usual interval formed by ignoring measurement error is approximately

$$I_2 = \{x: Y_* \in \hat{\alpha}_L + \hat{\beta}_L x \pm z_\alpha \hat{\sigma}_L\} .$$

This latter interval is strictly appropriate not for x_* but rather for $X_* = x_* + v$. The length of the confidence interval I_1 taking into account

measurement error in $\{x_i\}$ is, for large samples, proportional to

(A.2) $$L_1 \simeq 2\, z_\alpha\, \sigma_\epsilon/\beta$$

while that for the usual least squares analysis is proportional to

(A.3) $$L_2 \simeq 2\, z_\alpha(\sigma_\epsilon^2 + \beta^2\, \sigma_m^2)^{\frac{1}{2}}/\beta\ .$$

The ratio of these lengths is, noting (A.1),

(A.4) $$\frac{L_2}{L_1} \simeq (1 + \{\sigma_m/(\sigma_{\epsilon/\beta})\}^2)^{\frac{1}{2}}\ .$$

END

DTIC

7-86