

Applied and  
Computational  
Mathematics  
Division

NISTIR 89-4197

---

Center for Computing and Applied Mathematics

---

## *Orthogonal Distance Regression*

*Paul T. Boggs and Janet E. Rogers*

*November, 1989  
(Revised July, 1990)*

---

U.S. DEPARTMENT OF COMMERCE  
National Institute of Standards and Technology  
Gaithersburg, MD 20899

# Orthogonal Distance Regression <sup>\*</sup>

Paul T. Boggs <sup>†</sup>      Janet E. Rogers <sup>‡</sup>

## Abstract

Orthogonal Distance Regression (ODR) is the name given to the computational problem associated with finding the maximum likelihood estimators of parameters in measurement error models in the case of normally distributed errors. We examine the stable and efficient algorithm of Boggs, Byrd and Schnabel (*SIAM J. Sci. Stat. Comput.*, 8 (1987), pp. 1052–1078) for finding the solution of this problem when the underlying model is assumed to be nonlinear in both the independent variable and the parameters. We also describe the associated public domain software package, ODRPACK. We then review the results of a simulation study that compares ODR with ordinary least squares (OLS). We also present the new results of an extension to this study. Finally we discuss the use of the asymptotic covariance matrix for computing confidence regions and intervals for the estimated parameters. Our conclusions are that ODR is better than OLS for the criteria considered, and that ODRPACK can provide effective solutions and useful statistical information for nonlinear ODR problems.

## 1. Introduction

Much has been written about measurement error models and their properties. The definitive modern treatment of such problems, which are also known as errors in variables and generalized least squares, is given by Fuller [14]. Because

---

<sup>\*</sup>Contribution of the National Institute of Standards and Technology and not subject to copyright in the United States.

<sup>†</sup>Applied and Computational Mathematics Division, National Institute of Standards and Technology, Gaithersburg, MD 20899. INTERNET: boggs@cam.nist.gov

<sup>‡</sup>Applied and Computational Mathematics Division, National Institute of Standards and Technology, Boulder, CO, 80303-3328. INTERNET: jrogers@cam.nist.gov

of the complexity of these problems, most of the work in this area has concerned *linear* models. In this paper, however, we are primarily concerned with *nonlinear* measurement error models and the stable and efficient algorithm of Boggs, Byrd and Schnabel [5] for the numerical solution of the resulting nonlinear optimization problem that we refer to as the orthogonal distance regression problem. We also discuss the implementation of this algorithm in the software library ODRPACK and the use of this software to perform certain simulation studies that help to illuminate the differences between using an *orthogonal distance* criterion for estimating the parameters of a model and the use of the standard ordinary least squares (OLS) criterion. (See Boggs *et al.* [6], and Boggs, Rogers and Schnabel [8].) We conclude with another simulation study (Boggs and Rogers [7]) designed to test the effectiveness of the confidence intervals and regions obtained from the asymptotic covariance matrix.

The measurement error model can be defined in a number of ways. Here we assume that

$$(X_i, Y_i), \quad i = 1, \dots, n,$$

are observed random variables with underlying true values

$$(x_i, y_i), \quad i = 1, \dots, n,$$

where  $x_i \in \mathfrak{R}^m$  and  $y_i \in \mathfrak{R}$ . We also assume that  $\delta_i$  is the random error associated with  $x_i$ , that  $\epsilon_i$  is the random error associated with  $y_i$ , and that

$$\begin{aligned} X_i &= x_i - \delta_i \\ Y_i &= y_i - \epsilon_i. \end{aligned}$$

Finally, we assume that  $y_i$  is given as a function of  $x_i$  and a set of parameters  $\boldsymbol{\beta} \in \mathfrak{R}^p$ , i.e.,

$$y_i = f(x_i; \boldsymbol{\beta})$$

or

$$Y_i = f(X_i + \delta_i; \boldsymbol{\beta}) - \epsilon_i.$$

(Throughout this paper, we use bold face to denote the matrix or column vector whose components are the corresponding subscripted variables, e.g.,  $\boldsymbol{\beta} = (\beta_1, \beta_2, \dots, \beta_p)'$ , where  $'$  means transpose.) The function  $f$  may be either linear or nonlinear in its arguments, but must be a smooth function of these arguments. At this point, we make no restrictions on the distribution of  $x$ , i.e., we do not specify either a functional or structural model.

Any procedure that estimates  $\beta$  for a measurement error model should take into account that both  $X_i$  and  $Y_i$  have error. We do this by using an orthogonal distance to define the distance  $r_i$  from the point  $(X_i, Y_i)$  to the curve  $f(\tilde{x}; \tilde{\beta})$ , where  $\tilde{\cdot}$  denotes a non-distinguished value of the variable. That is, we take

$$r_i^2 = \min_{\tilde{\epsilon}_i, \tilde{\delta}_i} \{ \tilde{\epsilon}_i^2 + \tilde{\delta}_i^2 \}$$

subject to:  $Y_i = f(X_i + \tilde{\delta}_i; \tilde{\beta}) - \tilde{\epsilon}_i, \quad i = 1, \dots, n.$

Note that here, and in the following derivation, we assume that  $X_i$  is one dimensional to simplify the notation.

Under the assumptions that  $\epsilon_i \sim N(0, \sigma_{\epsilon_i}^2)$  and  $\delta_i \sim N(0, \sigma_{\delta_i}^2)$ , it follows that the maximum likelihood estimate  $\hat{\beta}$  is that which minimizes the sum of the squares of the  $r_i$ . Thus  $\hat{\beta}$  is the solution of

$$\min_{\tilde{\beta}, \tilde{\delta}, \tilde{\epsilon}} \sum_{i=1}^n \{ \tilde{\epsilon}_i^2 + \tilde{\delta}_i^2 \} \quad (1.1)$$

$$\text{subject to: } Y_i = f(X_i + \tilde{\delta}_i; \tilde{\beta}) - \tilde{\epsilon}_i, \quad i = 1, \dots, n. \quad (1.2)$$

Since the constraints (1.2) can be used to eliminate  $\tilde{\epsilon}$  from (1.1), this constrained problem is equivalent to the unconstrained minimization problem

$$\min_{\tilde{\beta}, \tilde{\delta}} \sum_{i=1}^n \left\{ [f(X_i + \tilde{\delta}_i; \tilde{\beta}) - Y_i]^2 + \tilde{\delta}_i^2 \right\}.$$

The final form of the optimization problem is obtained by allowing weights to be associated with the observations. We thus define the weighted *Orthogonal Distance Regression Problem* (ODR) to be

$$\min_{\tilde{\beta}, \tilde{\delta}} \sum_{i=1}^n w_i^2 \left\{ [f(X_i + \tilde{\delta}_i; \tilde{\beta}) - Y_i]^2 + d_i^2 \tilde{\delta}_i^2 \right\} \quad (1.3)$$

where  $w_i \geq 0$  and  $d_i > 0$ . The problem in this form can arise in statistical applications, and in curve fitting and other applications where the errors and their associated weights have no statistical connotation. In the remainder of this paper, however, we deal with problems where  $w_i = 1/\sigma_{\epsilon_i}$  and  $d_i = \sigma_{\epsilon_i}/\sigma_{\delta_i}$ .

In §2 we discuss the stable and efficient numerical algorithm in Boggs, Byrd and Schnabel [5] for solving this problem. This algorithm requires the same work

per iteration as the corresponding procedure for solving an OLS problem. The algorithm has been implemented in a Fortran subroutine library called ODRPACK (Boggs *et al.* [3]) that is publically available. The features of this package are also discussed.

In §3 we review the results of a simulation study (Boggs *et al.* [6]) that was designed to examine the differences between ODR and OLS solutions. This “baseline” study assumes that the ratios  $d_i = \sigma_{\epsilon_i}/\sigma_{\delta_i}$  are known and constant, and  $w_i = 1$  for all  $i$ . Under the chosen criteria ODR is shown to be superior to OLS. We then describe an extended study (Boggs, Rogers and Schnabel [8]) that relaxes the assumption that the  $d_i$  are known exactly. In this case, if the ratios are known to within a factor of 10, ODR is still preferred.

Fuller [14] shows how to derive the asymptotic covariance matrix for the ODR problem. In §4 we discuss the results of a further simulation study (Boggs and Rogers [7]) designed to ascertain the effectiveness of confidence regions and intervals constructed based on this matrix.

The results reported here allow the conclusion that ODR is preferred over OLS when the ratios  $d_i$  are reasonably well known. Furthermore, the existence of the algorithm and software described in §2 means that solving an ODR problem is just as easy as solving a nonlinear OLS problem, and that useful statistical information can be readily produced.

## 2. Algorithm and Software

The ODR problem (1.3) has been in the computing literature for several years, and several procedures for its solution have been proposed, cf., Britt and Luecke [9], Powell and MacDonald [15], Schnell [18], and Schwetlick and Tiller [19]. In this section, we review the recent stable and efficient algorithm of Boggs, Byrd and Schnabel [5], and its implementation in ODRPACK. (See Boggs *et al.* [3] and [4].) The algorithm is based on a *trust-region* strategy that yields a Levenberg-Marquardt type step. (See, e.g., Dennis and Schnabel [10].) The trust-region approach allows for a much more effective procedure for computing the so-called Levenberg-Marquardt parameter.

To describe the algorithm we again assume that  $X_i$  is one dimensional, which keeps the notation simpler. (The extension to  $m$  dimensions is straightforward, and ODRPACK allows this.) We first show that (1.3) can be viewed as an extended OLS problem. We then review the trust-region approach for solving nonlinear OLS problems, and show how the special structure of the extended problem

can be exploited to produce an efficient scheme. In fact the efficiency is comparable to that of a trust-region algorithm for solving the OLS problem obtained from the assumption that there are *no* errors in  $X_i$ . We conclude this section with a brief description of ODRPACK.

Let

$$\begin{aligned} g_i(\tilde{\boldsymbol{\beta}}, \tilde{\boldsymbol{\delta}}) &= w_i [f(X_i + \tilde{\delta}_i; \tilde{\boldsymbol{\beta}}) - Y_i] & i = 1, \dots, n \\ g_{i+n}(\tilde{\boldsymbol{\beta}}, \tilde{\boldsymbol{\delta}}) &= w_i d_i \tilde{\delta}_i & i = 1, \dots, n, \end{aligned}$$

and denote the 2-norm of  $\mathbf{g}$  by

$$\|\mathbf{g}(\tilde{\boldsymbol{\beta}}, \tilde{\boldsymbol{\delta}})\|^2 \equiv \sum_{i=1}^{2n} [g_i(\tilde{\boldsymbol{\beta}}, \tilde{\boldsymbol{\delta}})]^2.$$

Then (1.3) becomes

$$\min_{\tilde{\boldsymbol{\beta}}, \tilde{\boldsymbol{\delta}}} \|\mathbf{g}(\tilde{\boldsymbol{\beta}}, \tilde{\boldsymbol{\delta}})\|^2 \equiv \min_{\tilde{\boldsymbol{\beta}}, \tilde{\boldsymbol{\delta}}} \sum_{i=1}^{2n} [g_i(\tilde{\boldsymbol{\beta}}, \tilde{\boldsymbol{\delta}})]^2$$

which is an OLS problem with  $n + p$  parameters and  $2n$  observations.

Let

$$\tilde{\boldsymbol{\theta}} = \begin{pmatrix} \tilde{\boldsymbol{\beta}} \\ \tilde{\boldsymbol{\delta}} \end{pmatrix}$$

and denote by  $\mathbf{J} \in \Re^{2n \times (n+p)}$  the Jacobian of

$$\mathbf{g}(\tilde{\boldsymbol{\theta}}) = (g_1(\tilde{\boldsymbol{\theta}}), g_2(\tilde{\boldsymbol{\theta}}), \dots, g(\tilde{\boldsymbol{\theta}})_{2n})',$$

i.e.,

$$J_{ij} = \frac{\partial g_i}{\partial \tilde{\theta}_j}.$$

Then the *linearized model* at the current iterate,  $\boldsymbol{\theta}^c$ , given by the first two terms of the Taylor series, is

$$\mathbf{g}(\boldsymbol{\theta}^c) + \mathbf{J}(\boldsymbol{\theta}^c) \mathbf{s} \equiv \mathbf{g}^c + \mathbf{J}^c \mathbf{s} \quad (2.1)$$

where

$$\mathbf{s} = \tilde{\boldsymbol{\theta}} - \boldsymbol{\theta}^c.$$

(2.1) can then be used to compute a step,  $\mathbf{s}^c$ , from which the next iterate,  $\boldsymbol{\theta}^+$ , is found, i.e.,  $\boldsymbol{\theta}^+ = \boldsymbol{\theta}^c + \mathbf{s}^c$ . The obvious step that could be calculated from (2.1)

is the so-called Gauss-Newton step where  $\mathbf{s}^c$  is simply the (linear least squares) solution of

$$\min_{\mathbf{s}} \|\mathbf{g}^c + \mathbf{J}^c \mathbf{s}\|^2.$$

But even when the length of this step is controlled, e.g., by a line search, the resulting algorithm has not proven itself to be competitive with other procedures. (See, e.g., Dennis and Schnabel [10].) If, however, we explicitly recognize that the length of the step  $\mathbf{s}^c$  must be controlled by the extent to which (2.1) is a *good* model of  $\mathbf{g}$ , then a much better procedure arises.

Specifically, the *trust-region algorithm* chooses the current step  $\mathbf{s}^c$  as the solution to

$$\begin{aligned} \min_{\mathbf{s}} \|\mathbf{g}^c + \mathbf{J}^c \mathbf{s}\|^2 \\ \text{subject to: } \|\mathbf{s}\| \leq \tau \end{aligned} \quad (2.2)$$

where  $\tau$  is the *trust-region radius*. The value of  $\tau$  defines the region where (2.1) is a good model of  $\mathbf{g}$  and should not be thought of in a statistical sense. The value of  $\tau$  is automatically adjusted at each iteration based on whether  $\|\mathbf{g}(\boldsymbol{\theta}^c + \mathbf{s}^c)\| < \|\mathbf{g}^c\|$  and on a comparison of  $\|\mathbf{g}(\boldsymbol{\theta}^c + \mathbf{s}^c)\|$  with  $\|\mathbf{g}^c + \mathbf{J}^c \mathbf{s}^c\|$ . The actual tests employed in modern trust-region algorithms have been empirically determined over the years. (See Dennis and Schnabel [10] for a general discussion, and Boggs, Byrd and Schnabel [5] and Boggs *et al.* [3] for specific details.)

An outline of the trust-region algorithm is as follows.

Given an initial approximation  $\boldsymbol{\theta}^c$  and an estimate of the trust-region,  $\tau$ :

1. Solve (2.2) for  $\mathbf{s}^c$
2. Set  $\boldsymbol{\theta}^+ := \boldsymbol{\theta}^c + \mathbf{s}^c$
3. Test: If  $\|\mathbf{g}(\boldsymbol{\theta}^+)\| < \|\mathbf{g}^c\|$  then
  - set  $\boldsymbol{\theta}^c := \boldsymbol{\theta}^+$
  - adjust  $\tau$  if necessary
  - check convergence
 else
  - reduce  $\tau$
4. Go to Step 1

In the construction of portable software that is meant to be used for a variety of problems, it is common to solve (2.2) by first doing a QR factorization of the Jacobian matrix  $\mathbf{J}$ . (See, e.g., Dennis and Schnabel [10] or Stewart [20].) For a dense  $n \times p$  matrix the QR factorization requires  $O(np^2)$  operations. Since  $\mathbf{J}$  is  $2n \times (n + p)$ , it therefore requires  $O(n(n + p)^2) = O(n^3)$  operations to compute its QR factorization. The matrix  $\mathbf{J}$  has a special structure, however, that can be exploited to solve (2.2) more efficiently. In particular, let  $\mathbf{g}_1$  denote the first  $n$  components of  $\mathbf{g}$  and  $\mathbf{g}_2$  the last  $n$  components. Then

$$\begin{aligned} \mathbf{J} &= \begin{pmatrix} \frac{\partial \mathbf{g}_1}{\partial \boldsymbol{\beta}} & \frac{\partial \mathbf{g}_1}{\partial \tilde{\boldsymbol{\delta}}} \\ \frac{\partial \mathbf{g}_2}{\partial \boldsymbol{\beta}} & \frac{\partial \mathbf{g}_2}{\partial \tilde{\boldsymbol{\delta}}} \end{pmatrix} \\ &= \begin{pmatrix} \frac{\partial \mathbf{g}_1}{\partial \boldsymbol{\beta}} & \mathbf{H} \\ \mathbf{0} & \mathbf{D} \end{pmatrix} \end{aligned}$$

where

$$\mathbf{H} = \text{diag} \left\{ w_i \frac{\partial f(X_i + \tilde{\boldsymbol{\delta}}_i; \tilde{\boldsymbol{\beta}})}{\partial \tilde{\boldsymbol{\delta}}_i}, i = 1, \dots, n \right\}$$

and

$$\mathbf{D} = \text{diag} \{ w_i d_i, i = 1, \dots, n \}. \quad (2.3)$$

By exploiting the structure provided by the existence of the zero block and the two diagonal blocks,  $\mathbf{H}$  and  $\mathbf{D}$ , the amount of work per iteration required to solve an ODR problem can be reduced from  $O(n^3)$  to  $O(np^2)$  operations plus, on average, less than two evaluations of the model function  $f$ . Since there are typically many more observations than parameters, this difference can have a profound influence on the size of problem that can be practicably solved. We point out that if the errors in  $X_i$  are ignored and the resulting OLS problem is solved, the amount of work per iteration required is also  $O(np^2)$  operations plus an average of less than two evaluations of  $f$ . Using our algorithm, therefore, an ODR problem can be solved as efficiently as an OLS problem. See Boggs *et al.* [3] for the details.

We conclude this section with a brief description of the Fortran subroutine library, ODRPACK, that we have written to implement the ODR algorithm described above. A complete description is contained in Boggs *et al.* ([3] and [4]). The package is portable and has been successfully run on many machines from PCs to supercomputers.

ODRPACK is designed to solve both OLS and ODR problems, handling many levels of user sophistication and problem difficulty.

- It is easy to use, providing two levels of user control of the computations, extensive error handling facilities, optional printed reports summarizing the iterations and/or the final results, and no size restrictions other than effective machine size.
- The necessary Jacobian matrices are approximated numerically if they are not supplied by the user. The correctness of user supplied derivatives can also be verified by the derivative checking procedure provided.
- Both weighted and unweighted analysis can be performed. Unequal weights  $w_i$ ,  $i = 1, \dots, n$ , are allowed, and a different value of  $d_{ik}$ ,  $i = 1, \dots, n$  and  $k = 1, \dots, m$ , can be specified for each component of  $\mathbf{X}$ . A feature has also been incorporated that allows the same weight to apply to many components without having to explicitly enter it more than once.
- Subsets of the unknowns can be treated as constants with their values held fixed at their input values, allowing the user to examine the results obtained by estimating subsets of the unknowns of a general model without rewriting the code that computes  $f$ .
- The covariance matrix and the standard errors for the model parameter estimators are optionally provided. (See §4.)
- ODRPACK automatically compensates for problems in which the model parameters and/or unknown errors in the independent variables vary widely in magnitude.

A copy of ODRPACK may be obtained free of charge from the authors.

### 3. Simulation Studies

Researchers including Fuller [14] and Reilman *et al.* [17] provide results that give theoretical conditions for choosing ODR over OLS and vice versa. These results, however, are only for straight line models, and, to our knowledge, there are no similar results for either general linear models or for nonlinear models. This state of affairs provides the motivation for a baseline simulation study that compares the

performance of ODR and OLS on nonlinear models. These results are contained in Boggs *et al.* [6].

The central assumption in Boggs *et al.* [6] is that the ratios

$$d_i = \sigma_{\epsilon_i} / \sigma_{\delta_i}$$

are known exactly. We first review this study, and then present results that extend this work to the case when the  $d_i$  are not known exactly.

The existence of ODRPACK as described in §2 not only makes these studies possible, but also makes them of practical interest. Since this code is publically available, other researchers and practitioners can solve their own ODR problems and conduct similar simulation studies on other models.

The study in Boggs *et al.* [6] is the first to consider models that are nonlinear in both  $\beta$  and  $x$ , although studies involving linear and quadratic models have been reported in Amemiya and Fuller [1], Schnell [18], and Wolter and Fuller [21], for example. Our procedure is to compare the bias, variance, and mean square error of parameter estimates and function estimates obtained using both ODR and OLS. We use four forms of the model function  $f$  with two parameter sets per function. We generate errors for each using  $d_i = d^*$  for seven values of  $d^*$  ranging from 0.1 to  $\infty$ , replicating each problem configuration 500 times. Each of the resulting 91,000 problems is then solved using ODR and OLS. (Recall that the correct values of  $d_i$  are used to obtain the ODR solutions.) Both the ODR and OLS computations are done by ODRPACK.

To be more specific, we consider a straight line model, a quadratic model, an exponential model, and a sine model. The two true parameter sets for each of these models are chosen so that the maximum slope of  $f$  over the range of the  $X$  data is 1 and 10, respectively. The  $x_i$  are taken as 51 equally spaced points on  $[-1,1]$ ; thus this study is concerned with a functional model. The errors  $\delta$  and  $\epsilon$  are generated as normally distributed and adjusted so that the expected sum of the squared errors is constant over the seven values of  $d^*$ . The values of  $d^*$  that we consider are 0.1, 0.5, 1.0, 2.0, 10.0, 100.0, and  $\infty$ , where  $d^* = \infty$  indicates that there are no errors in the values  $X_i$ .

The main conclusion of this study is that ODR is *better* than OLS for all of the criteria that we consider.

- For parameter bias, ODR is no worse than OLS in 98% of the cases studied and is a clear winner 50% of the time.
- For parameter variance and mean square error, ODR is no worse in 98% of the cases and is appreciably better in 23%.

- Similar results are obtained for the function estimates.
- For the straight line model, we do not see a preference for OLS over ODR, as is predicted by Reilman *et al.* [17]. ODR produces a significantly lower bias than OLS, but the variance and mean square error values are only slightly smaller for OLS. We thus conclude that ODR is still preferable in this case.
- In accordance with the observations of Wolter and Fuller [21] and Anderson [2] our computational results do not reflect the infinite moments of ODR.

In practice, the ratios  $d_i$  are seldom known exactly. Thus it is of interest to extend the above results to the case where an approximate value of each  $d_i$  is used in the ODR procedure. The results of such a study are presented next.

To make this second Monte Carlo study tractable, we reduce its scope somewhat. We consider the same seven functions as described above, but with only one parameter set per function. For this study, the ratios used to compute the generated errors are  $d^*$  equal to 0.1, 1.0, and 10.0, and the values of the error ratios that are used by ODRPACK in estimating the parameters are  $d_i = \gamma d^*$ , where  $\gamma$  has the values 0.1, 0.5, 1.0, 2.0 and 10. That is, we investigate cases where the error ratio used in the estimation procedure is correct to within a factor of 10. We again compare the ODR results to an OLS fit using the same criteria as in the baseline study reported above.

We give a complete description of the study and the results in Boggs, Rogers and Schnabel [8]. In summary, when the  $d_i$  are correct to within a factor of 10.0, then:

- The bias of the parameter estimates using ODR is no worse than the bias using OLS in 88% of the cases studied, and is significantly smaller 59% of the time.
- The variance of the parameter estimates using ODR is no worse in 90% of the cases and is smaller in 40%.
- The mean square error of the parameter estimates using ODR is no worse in 87% of the cases and is smaller in 50%.

Similar results are observed for the function estimates. In addition, there appears to be some support for overestimating rather than underestimating the  $d_i$ .

We conclude from these two studies that if the  $d_i$  are known to within a factor of 10, then ODR is preferable to OLS. In the next section, however, we describe

how estimated confidence regions may break down if the  $d_i$  are not known to within a factor of 2.

#### 4. The Asymptotic Covariance Matrix

In many parameter estimation problems, confidence regions and/or confidence intervals for the estimated parameters are needed. In the linear OLS case, these can be obtained exactly. In the nonlinear OLS case and in ODR, the computation of exact confidence regions and intervals is computationally intractable, and so approximate methods must be used.

Fuller [14] derives the asymptotic form of the covariance matrix for the estimated parameters of measurement error models. This can be used to construct approximate confidence regions and intervals. In Boggs and Rogers [7] we discuss the efficient computation of this covariance matrix in the context of the algorithm and software of §2. Using a Monte Carlo simulation study, we then assess the quality of confidence regions and intervals computed from this covariance matrix. In this section we review that work.

The asymptotic covariance matrix can be defined using the extended OLS problem of §2, namely,

$$\min_{\tilde{\boldsymbol{\theta}}} \sum_{i=1}^{2n} g_i(\tilde{\boldsymbol{\theta}})^2 \quad (4.1)$$

and the corresponding Jacobian matrix  $\mathbf{J} \in \Re^{2n \times (n+p)}$  with  $(i, j)$  element defined by

$$J_{ij} = \frac{\partial g_i}{\partial \tilde{\theta}_j}.$$

Let  $\hat{\boldsymbol{\theta}}$  be the estimate of  $\boldsymbol{\theta}$  from (4.1) and let

$$\hat{\sigma}^2 = \frac{\mathbf{g}(\hat{\boldsymbol{\theta}})' \boldsymbol{\Omega} \mathbf{g}(\hat{\boldsymbol{\theta}})}{n + p}$$

where

$$\boldsymbol{\Omega} = \begin{pmatrix} \mathbf{W}^2 & \mathbf{0} \\ \mathbf{0} & \mathbf{D}^2 \end{pmatrix},$$

$$\mathbf{W} = \text{diag}\{w_i, i = 1, \dots, n\},$$

and  $\mathbf{D}$  is given by (2.3). Also, assume that  $\mathbf{\Omega}$  correctly reflects the distribution of the errors  $\boldsymbol{\epsilon}$  and  $\boldsymbol{\delta}$  as specified in §1. Then the covariance matrix is

$$\hat{\mathbf{V}} = \hat{\sigma}^2 [\mathbf{J}'\mathbf{\Omega}\mathbf{J}]^{-1},$$

which Fuller [14] proves is asymptotically correct as the error variance tends to zero.

The matrix  $\hat{\mathbf{V}}$  can be used to construct approximate confidence regions and intervals for the estimated parameters  $\hat{\boldsymbol{\theta}}$ . We call this method the *linearization method* since it is based on the assumption that  $\mathbf{g}$  is adequately approximated by a linear function of the form (2.1) in the neighborhood of  $\hat{\boldsymbol{\theta}}$ . The goodness of the resulting *linearized regions* and *linearized intervals* depends on the nonlinearity of  $f$  and the size of  $\hat{\sigma}^2$ . Donaldson and Schnabel [11] show that the linearized confidence intervals for nonlinear OLS problems appear to be reasonable in practice, but that the linearized confidence regions can be inadequate. (More accurate, but comensurately more expensive approximations can be obtained, as described by, e.g., Donaldson and Schnabel [11] and Efron [13].)

To assess linearized confidence regions and intervals in the context of ODR problems, we conduct a Monte Carlo simulation similar to that in Donaldson and Schnabel [11]. Our study consists of four example problems. Three of these problems exist in the literature. They are Fuller [14, Example 3.2.2], Ratkowsky [16, Example 6.11], and Draper and Smith [12, Chapter 10, Problem E]. The fourth example is taken from a consulting session involving a “psychophysical” experiment evaluating the ability of human subjects to perceive a visual signal as a function of the intensity of the signal.

For each example we generate 500 sets of “observed” data with normal random error and true error ratios  $d_i^*$ . We then solve each of the resulting problems using ODRPACK. As in the second study reported in §3, this study also solves each problem using a range of assumed error ratios  $d_i = \gamma d_i^*$ , where  $\gamma = 0.1, 0.5, 1.0, 2.0,$  and  $10.0$ . For each of the 500 data sets, we also construct the 95% confidence regions and intervals for the estimate  $\hat{\boldsymbol{\theta}}$  using the linearization method. Finally, we calculate the *observed coverage*, which is the percentage of cases in which the true value is actually contained in the computed region or interval. Our results, presented in Boggs and Rogers [7], are summarized as follows:

- When the  $d_i$  are correct, we obtain surprisingly good estimates of the confidence regions and intervals as compared to the results of Donaldson and Schnabel [11], who often found the observed coverage for a nominal 95%

confidence region to be less than 80%. We assume that this is due to our choice of examples and is not a general property of ODR.

- The confidence region estimates significantly degrade, even when the estimated values of  $d_i$  are off by no more than a factor of 2.
- The confidence intervals when the  $d_i$  are correct are also quite good and remain good when the  $d_i$  are only off by a factor of 2.
- The confidence intervals significantly degrade when the  $d_i$  are off by a factor of 10.
- Confidence regions and intervals are better for  $\hat{\beta}$  than for  $\hat{\delta}$ .
- In the case of confidence intervals for  $\hat{\beta}$ , overestimation of the  $d_i$  is preferable to underestimation. For  $\hat{\delta}$  the reverse is true. This occurs because, as the  $d_i$  are increased,  $\hat{\delta}$  becomes more restricted, whereas as the  $d_i$  are decreased, an artificially small value of  $\hat{\sigma}$  is produced.

Note that  $\hat{\mathbf{V}}$  is the covariance matrix for the entire parameter vector  $\hat{\boldsymbol{\theta}}$ . In practice, however, one is typically only interested in the covariance matrix for the model parameters  $\hat{\boldsymbol{\beta}}$ . This is the upper left  $p \times p$  part of  $\hat{\mathbf{V}}$ , which we denote  $\hat{\mathbf{V}}_{\beta}$ . Just as the special structure of  $\mathbf{J}$  allows us to create a fast algorithm for solving (2.2), that same structure permits us to equip ODRPACK with an efficient and stable means of calculating  $\hat{\mathbf{V}}_{\beta}$ . The formulas are contained in Boggs and Rogers [7] and are based on the work in Boggs, Byrd and Schnabel [5].

The overall conclusion from this third study is that the linearized confidence regions and especially confidence intervals have some validity for ODR problems when the  $d_i$  are known to within a factor of 2. Furthermore, the computations needed to obtain these linearized regions and intervals for  $\hat{\boldsymbol{\beta}}$  are cheap, and are automatically produced by ODRPACK. Thus we advocate the use of linearized confidence regions and intervals for measurement error models in the same spirit, and with the same caveats, that accompany their use in nonlinear OLS problems.

**Acknowledgement.** The authors gratefully acknowledge our colleagues R. H. Byrd, R. B. Schnabel, and C. H. Spiegelman for their part in the work that is reported here. We also thank K. R. Eberhardt, H. K. Iyer and G. W. Stewart for many useful discussions on this topic.

## References

- [1] Y. AMEMIYA AND W. A. FULLER, *Estimation for the nonlinear functional relationship: errors-in-variables model*, Report to the U. S. Bureau of the Census on J.S.A. 85-1, Iowa State University, Ames, Iowa, 1985.
- [2] T. W. ANDERSON, *Estimation of linear functional relationships: approximate distributions and connections with simultaneous equations in econometrics*, Journal of the Royal Statistical Society (Series B), 38 (1976), pp. 1–36.
- [3] P. T. BOGGS, R. H. BYRD, J. R. DONALDSON, AND R. B. SCHNABEL, *ODR-PACK — software for weighted orthogonal distance regression*, ACM Transactions on Mathematical Software, 15 (1989), pp. 348–364.
- [4] ———, *User's reference guide for ODRPACK — software for weighted orthogonal distance regression*, Internal Report 89-4103, National Institute of Standards and Technology, Gaithersburg, Maryland, 1989.
- [5] P. T. BOGGS, R. H. BYRD, AND R. B. SCHNABEL, *A stable and efficient algorithm for nonlinear orthogonal distance regression*, SIAM Journal of Scientific and Statistical Computing, 8 (1987), pp. 1052–1078.
- [6] P. T. BOGGS, J. R. DONALDSON, R. B. SCHNABEL, AND C. H. SPIEGELMAN, *A computational examination of orthogonal distance regression*, Journal of Econometrics, 38 (1988), pp. 169–201.
- [7] P. T. BOGGS AND J. E. ROGERS, *The computation and use of the asymptotic covariance matrix for measurement error models*, Internal Report 89-4102 (Revised 1990), National Institute of Standards and Technology, Gaithersburg, Maryland, 1989.
- [8] P. T. BOGGS, J. E. ROGERS, AND R. B. SCHNABEL, *A second Monte Carlo study of weighted orthogonal distance regression*, Internal Report (in progress), National Institute of Standards and Technology, Gaithersburg, Maryland, 1991.
- [9] H. L. BRITT AND R. H. LUECKE, *The estimation of parameters in nonlinear implicit models*, Technometrics, 15 (1973), pp. 233–247.
- [10] J. E. DENNIS AND R. B. SCHNABEL, *Numerical Methods for Unconstrained Optimization and Nonlinear Equations*, Prentice-Hall, Englewood Cliffs, New Jersey, 1983.

- 
- [11] J. R. DONALDSON AND R. B. SCHNABEL, *Computational experience with confidence regions and confidence intervals for nonlinear least squares*, *Technometrics*, 29 (1987), pp. 67–82.
- [12] N. R. DRAPER AND H. SMITH, *Applied Regression Analysis, Second Edition*, John Wiley and Sons, New York, New York, 1981.
- [13] B. EFRON, *The Jackknife, the Bootstrap and Other Resampling Plans*, Society for Industrial and Applied Mathematics, Philadelphia, Pennsylvania, 1985.
- [14] W. A. FULLER, *Measurement Error Models*, John Wiley and Sons, New York, New York, 1987.
- [15] D. R. POWELL AND J. R. MACDONALD, *A rapidly convergent iterative method for the solution of the generalized nonlinear least squares problem*, *Computer Journal*, 15 (1972), pp. 148–155.
- [16] D. A. RATKOWSKY, *Nonlinear Regression Modeling*, Marcel Dekker, New York, New York, 1983.
- [17] M. A. REILMAN, R. F. GUNST, AND M. Y. LAKSHMINARAYANAN, *Stochastic regression with errors in both variables*, *Journal of Quality Technology*, 18 (1986), pp. 162–169.
- [18] D. SCHNELL, *Maximum likelihood estimation of parameters in the implicit nonlinear errors-in-variables model*, Master's thesis, Iowa State University Department of Statistics, Ames, Iowa, 1983.
- [19] H. SCHWETLICK AND V. TILLER, *Numerical methods for estimating parameters in nonlinear models with errors in the variables*, *Technometrics*, 27 (1985), pp. 17–24.
- [20] G. W. STEWART, *Introduction to Matrix Computations*, Academic Press, New York, New York, 1973.
- [21] K. M. WOLTER AND W. A. FULLER, *Estimation of nonlinear errors-in-variables models*, *The Annals of Statistics*, 10 (1982), pp. 539–548.