Bulletin of the Seismological Society of America, Vol. 83, No. 2, pp. 469–487, April 1993

METHODS FOR REGRESSION ANALYSIS OF STRONG-MOTION DATA

BY WILLIAM B. JOYNER AND DAVID M. BOORE

ABSTRACT

We introduce a new computational method for implementing Brillinger and Preisler's (1984, 1985) one-stage maximum-likelihood analysis of strong-motion data. We also reexamine two-stage methods and agree with Masuda and Ohtake (1992) that rigorous analysis requires off-diagonal terms in the weighting matrix for the second-stage regression but note that Masuda and Ohtake failed to account for the earthquake-to-earthquake component of variance. Analysis by Monte Carlo methods shows that both one-stage and two-stage methods, properly applied, are unbiased and that they have comparable uncertainties. Both give the same correct results when applied to the data that Fukushima and Tanaka (1990) have shown cannot be satisfactorily analyzed by ordinary least squares. The two-stage method is more efficient computationally, but for typical problems neither method requires enough time to make efficiency important. Of the two methods, only the two-stage method can readily be used with the techniques described by Toro (1981) and McLaughlin (1991) for overcoming the bias due to instruments that do not trigger.

INTRODUCTION

Empirical equations for predicting strong ground motion are typically fit to the strong-motion data set by the method of ordinary least squares. Campbell (1981, 1989) used weighted least squares in an attempt to compensate for the nonuniform distribution of data with respect to distance. We introduced a two-stage regression method designed to decouple the determination of the magnitude dependence from the determination of the distance dependence (Joyner and Boore, 1981). In the first stage, the distance dependence was determined along with a set of amplitude factors, one for each earthquake. In the second stage, the amplitude factors were regressed against magnitude to determine the magnitude dependence. Fukushima and Tanaka (1990) used a similar two-stage method on the Japanese peak horizontal acceleration data set and compared results with those from one-stage ordinary least squares. They showed that the one-stage ordinary least-squares results were seriously in error. They attributed the error to the strong correlation between magnitude and distance and the resulting trade-off between magnitude dependence and distance dependence. The correct distance dependence, given by the two-stage method and verified by analyzing individual earthquakes separately, showed a much stronger decay of peak acceleration with distance than the one-stage ordinary least-squares method, which had been used previously.

In our original use of the two-stage method (Joyner and Boore, 1981, 1982), we included in the second-stage regression for peak acceleration only those earthquakes that had been recorded at more than one station, and we gave equal weight to each earthquake included. For peak velocity and response spectra, there were so few earthquakes in the data set that we were compelled

to use them all, and we gave each equal weight. Later we proposed a diagonal weighting scheme to be used in the second-stage regression (Joyner and Boore, 1988). Fukushima and Tanaka's (1990) procedure for the second-stage regression had the effect of weighting each earthquake by the number of recordings. Masuda and Ohtake (1992) proposed a weighting matrix for the second-stage regression different from any used earlier. They showed that off-diagonal terms need to be included in the weighting matrix, because the amplitude factors that are the dependent variables in the second-stage regression are mutually correlated as a consequence of the fact that they were determined in the first-stage regression along with the parameters that control the distance dependence. As Fukushima and Tanaka (1992) point out, however, the off-diagonal terms are small in magnitude.

Brillinger and Preisler (1984, 1985) introduced what they called the random-effects model, which incorporated an explicit earthquake-to-earthquake component of variance in addition to the record-to-record component. They described one-stage maximum-likelihood methods for evaluating the parameters in the prediction equation. Abrahamson and Youngs (1992) introduced an alternative algorithm, which they considered more stable though less efficient.

The concept of an earthquake-to-earthquake component of variance is implicit in the two-stage regression methods. The two-stage methods are not, however, exactly equivalent to the one-stage maximum-likelihood methods, and the relationship of one to the other is not obvious. Both the one-stage and two-stage methods are based on maximum likelihood. In the one-stage methods, the parameters are all determined simultaneously by maximizing the likelihood of the set of observations. In the two-stage methods, the parameters controlling distance dependence and a set of amplitude factors, one for each earthquake, are determined in the first stage, by maximizing the likelihood of the set of observations. The parameters controlling magnitude dependence are then determined in the second stage by maximizing the likelihood of the set of amplitude factors. On the face of it, one might expect either method to give satisfactory results. The one-stage method may be more elegant mathematically, but the two-stage method is conceptually simpler. The two-stage method can be considered the analytical equivalent of the graphical method employed by Richter (1935, 1958) in developing the attenuation curve that forms the basis for the local magnitude scale in southern California. As we will show, the two-stage method is more efficient computationally, but the one-stage method requires so little computer time that efficiency is not really an issue.

In view of the number of different approaches that have been proposed, we believe it timely to attempt to sort out how these approaches relate to each other. We begin by developing our own computational method for one-stage maximum-likelihood analysis, which makes use of the conventional mathematics of regression analysis. We then reexamine two-stage methods and derive the correct weighting for the second stage. Finally, we examine estimation errors and compare the one-stage method with the two-stage method by Monte Carlo simulations. To illustrate the discussion, we use our original peak horizontal acceleration data set (Joyner and Boore, 1981), which differs slightly from the data set used later (Joyner and Boore, 1982, 1988). This choice facilitates comparison with Brillinger and Preisler (1984, 1985), who used our original data set, which includes 182 records from 23 earthquakes.

One-Stage Maximum-Likelihood Methods

We use the formulation of Brillinger and Preisler (1984, 1985) and fit the data by the equation

$$\log A_n = a + b(M_n - 6) - \log\left(d_n^2 + h^2\right)^{1/2} + c\left(d_n^2 + h^2\right)^{1/2} + \epsilon_r + \epsilon_e, \quad (1)$$

where A_n is peak horizontal acceleration for the nth record; M_n is the moment magnitude (Hanks and Kanamori, 1979) of the earthquake corresponding to the nth record; d_n is the shortest distance from the recording site of the nth record to the vertical projection of the earthquake fault rupture on the surface of the Earth; ϵ_r is an independent random variable that takes on a specific value for each record; ϵ_e is an independent random variable that takes on a specific value for each earthquake; and a, b, h, and c are parameters to be determined. The records are numbered so that all records from the same earthquake are consecutive. For all numerical results, the units of acceleration and distance are g and km, respectively. The mean of ϵ_e over the population of earthquakes is zero, and the variance is σ_e^2; ϵ_e represents the earthquake-to-earthquake component of variability. The mean of ϵ_r over the population of records is zero, and the variance is σ_r^2; ϵ_r represents the site-to-site component of variability plus all other sources of variability not represented by ϵ_e.

Strictly speaking, the site-to-site component of variability should be separated out in the same way as the earthquake-to-earthquake component. ϵ_r should be considered the sum of two variables: ϵ_s, an independent random variable that takes on a specific value for each site, and ϵ_o, an independent random variable that takes on a specific value for each record. Because there are few records in the data set corresponding to different earthquakes recorded at the same site, however, we are entitled to take advantage of the considerable simplification afforded by lumping ϵ_s and ϵ_o. The analysis for separate ϵ_s and ϵ_o is given in Appendix A.

We propose our own method for determining the parameters. Equation (1) is nonlinear in the parameters because of the terms involving h. We linearize the problem with a Taylor's-series expansion about trial values of the parameters (Draper and Smith, 1981, pp. 462–464) and set

$$\mathbf{Y} = \begin{bmatrix} \log A_1 + \log\left(d_1^2 + h'^2\right)^{1/2} \\ \log A_2 + \log\left(d_2^2 + h'^2\right)^{1/2} \\ \vdots \\ \log A_N + \log\left(d_N^2 + h'^2\right)^{1/2} \end{bmatrix}, \quad (2)$$

$$\mathbf{B} = \begin{bmatrix} a \\ b \\ c \\ \Delta h \end{bmatrix}, \quad (3)$$

and

$$\mathbf{X} = \begin{bmatrix} 1 & M_1 - 6 & (d_1^2 + h'^2)^{1/2} & \left\{\frac{\partial}{\partial h}\left[c'(d_1^2 + h^2)^{1/2} - \log(d_1^2 + h^2)^{1/2}\right]\right\}_{h=h'} \\ 1 & M_2 - 6 & (d_2^2 + h'^2)^{1/2} & \left\{\frac{\partial}{\partial h}\left[c'(d_2^2 + h^2)^{1/2} - \log(d_2^2 + h^2)^{1/2}\right]\right\}_{h=h'} \\ \vdots & \vdots & \vdots & \vdots \\ 1 & M_N - 6 & (d_N^2 + h'^2)^{1/2} & \left\{\frac{\partial}{\partial h}\left[c'(d_N^2 + h^2)^{1/2} - \log(d_N^2 + h^2)^{1/2}\right]\right\}_{h=h'} \end{bmatrix}, \quad (4)$$

where N is the total number of data points and h' and c' are trial values of h and c. In each new iteration $h' + \Delta h$ replaces h' of the previous iteration. Practically any positive initial value should work for h'. We use 1.0 km. Zero, however, will not work, because the partial derivatives that form the last column of $\mathbf{X}$ are zero for zero h'. Equation (1) can now be replaced by the system

$$\mathbf{Y} = \mathbf{XB} + \mathbf{e}, \quad (5)$$

where $\mathbf{e}$ is the vector of deviations, which are composed of ϵ_r and ϵ_e. Equation (5) is linear in the parameters to be determined and will be applied in an iterative scheme. We assume that the components of $\mathbf{e}$ are normally distributed with zero mean and variance-covariance matrix $\mathbf{V}$. Ordinary least squares could be used to solve equation (5) if the variance-covariance matrix $\mathbf{V}$ were diagonal with equal elements (Searle, 1971, p. 87), but $\mathbf{V}$ is not diagonal because there is a correlation between values recorded in the same earthquake. Generalized least squares (Searle, 1971, p. 87), which involves weighting by the inverse of the matrix $\mathbf{V}$, could be used if $\mathbf{V}$ were known in advance, but, as will be shown, $\mathbf{V}$ depends on σ_r and σ_e, and so we turn to the maximum-likelihood method. The likelihood of the sample of observations is

$$L = (2\pi)^{-N/2}|\mathbf{V}|^{-1/2}\exp\left[-\tfrac{1}{2}(\mathbf{Y} - \mathbf{XB})^T\mathbf{V}^{-1}(\mathbf{Y} - \mathbf{XB})\right] \quad (6)$$

(Searle, 1971, p. 87), where ^{T} denotes matrix transposition and $||$ denotes the determinant. For a given $\mathbf{V}$, maximizing L with respect to $\mathbf{B}$ is the equivalent of minimizing

$$(\mathbf{Y} - \mathbf{XB})^T\mathbf{V}^{-1}(\mathbf{Y} - \mathbf{XB}). \quad (7)$$

The solution (Searle 1971, p. 87) is

$$\hat{\mathbf{B}} = (\mathbf{X}^T\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}^T\mathbf{V}^{-1}\mathbf{Y}. \quad (8)$$

To derive an expression for the variance-covariance matrix $\mathbf{V}$, we return to equation (1) and note that a component of $\mathbf{e}$ represents the sum $\epsilon_r + \epsilon_e$ and that ϵ_r takes on a specific value for each record and ϵ_e takes on a specific value for each earthquake. The covariance of two components of $\mathbf{e}$ corresponding to different earthquakes is zero. The covariance of two components corresponding to the same earthquake is σ_e^2, the variance of ϵ_e. The variance of an individual component is $\sigma_r^2 + \sigma_e^2$, the variance of $\epsilon_r + \epsilon_e$. Recall that the records are numbered so that all records from the same earthquake are consecutive. The

variance-covariance matrix $\mathbf{V}$ is therefore block-diagonal with each block corresponding to an earthquake. We define a normalized matrix $\mathbf{v}$ by the equation

$$\sigma^2\mathbf{v} = \mathbf{V}, \tag{9}$$

where $\sigma^2 = \sigma_r^2 + \sigma_e^2$. The matrix $\mathbf{v}$ is also block-diagonal

$$\mathbf{v} = \begin{bmatrix} \mathbf{v}_1 & 0 & \cdots & 0 \\ 0 & \mathbf{v}_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \mathbf{v}_{Ne} \end{bmatrix}, \tag{10}$$

where N_e is the number of earthquakes. The submatrix $\mathbf{v_i}$ corresponding to earthquake i is given by

$$\mathbf{v_i} = \begin{bmatrix} 1 & \gamma & \cdots & \gamma \\ \gamma & 1 & \cdots & \gamma \\ \vdots & \vdots & \ddots & \vdots \\ \gamma & \gamma & \cdots & 1 \end{bmatrix}, \tag{11}$$

where $\gamma = \sigma_e^2/(\sigma_r^2 + \sigma_e^2)$. The rank of matrix $\mathbf{v_i}$ is R_i, the number of recordings for earthquake i. Substituting equation number (9) into equation (8) gives

$$\hat{\mathbf{B}} = (\mathbf{X}^T\mathbf{v}^{-1}\mathbf{X})^{-1}\mathbf{X}^T\mathbf{v}^{-1}\mathbf{Y}. \tag{12}$$

Substituting equation number (9) into equation (6) gives for the likelihood

$$L = (2\pi)^{-N/2}|\sigma^2\mathbf{v}|^{-1/2}\exp\left[-\tfrac{1}{2}(\mathbf{Y} - \mathbf{XB})^T(\sigma^2\mathbf{v})^{-1}(\mathbf{Y} - \mathbf{XB})\right]. \tag{13}$$

Taking the natural logarithm of equation (13) gives

$$\ln L = -\frac{N}{2}\ln(2\pi) - \frac{N}{2}\ln(\sigma^2) - \frac{1}{2}\ln|\mathbf{v}| - \frac{1}{2}(\mathbf{Y} - \mathbf{XB})^T\mathbf{v}^{-1}(\mathbf{Y} - \mathbf{XB})/\sigma^2. \tag{14}$$

The likelihood L must be maximized over all γ, $\mathbf{B}$, h, and σ^2. For fixed γ, L can be maximized with respect to $\mathbf{B}$ and h by iterating on equation (12) until $|\Delta h/h|$ is reduced below a specified limit, generally 10^{-3}. Since $\hat{\mathbf{B}}$ and h do not depend on σ^2 for fixed γ, we may then proceed to maximize L with respect to σ^2 without recomputing $\hat{\mathbf{B}}$ or h; we differentiate equation (14) with respect to σ^2, set the result equal to zero, and solve for σ^2. The solution is

$$\sigma^2 = (\mathbf{Y} - \mathbf{X}\hat{\mathbf{B}})^T\mathbf{v}^{-1}(\mathbf{Y} - \mathbf{X}\hat{\mathbf{B}})/N. \tag{15}$$

For each value of γ, we compute values of $\hat{\mathbf{B}}$, h, σ^2, and the likelihood L maximized with respect to $\hat{\mathbf{B}}$, h, and σ^2. The final solution corresponds to the value of γ for which the logarithm of the likelihood (equation 14) is maximum. The solution is found numerically by searching over γ using the search routine GOLDEN given by Press *et al.* (1989, p. 282).

The value of σ^2 given by equation (15) is not unbiased. An unbiased estimate is

$$\hat{\sigma}^2 = (\mathbf{Y} - \mathbf{X}\hat{\mathbf{B}})^T \mathbf{v}^{-1}(\mathbf{Y} - \mathbf{X}\hat{\mathbf{B}})/(N - 4), \tag{16}$$

where $N - 4$ is the number of degrees of freedom, 4 being the rank of the matrix $\mathbf{X}$. This result is obtained by noting that, for the final value of γ, the determination of $\hat{\mathbf{B}}$ is a weighted least-squares problem with weighting matrix $\mathbf{v}^{-1}$. Draper and Smith (1981, pp. 108–109) have shown that such a problem is the equivalent of an ordinary least-squares problem in a transformed variable. Equation (16) follows from applying the usual rules (Searle, 1971, p. 93) to the equivalent ordinary least-squares problem.

The rank of the matrix $\mathbf{v}$ is equal to the total number of recordings in the data set, so it is advantageous to determine the inverse and determinant analytically. The formulas for doing so are given in Appendix B.

The coefficients of equation (1) determined by the method just described are compared in Table 1 with the values obtained by Brillinger and Preisler (1985) for our original data set (Joyner and Boore, 1981). Because the problem is nonlinear, we use Monte Carlo methods to estimate uncertainties in the parameters. The results are given in a subsequent section. In the first row of Table 1, $(a - 6b)$ is compared, rather than a, because Brillinger and Preisler wrote their equation in terms of M rather than $M - 6$. The agreement is nearly perfect. The calculations that produced our result in Table 1 took 62 sec of CPU time on a VAXstation 3100 for a data set consisting of 182 records from 23 earthquakes.

Two-Stage Methods

Return to equation (1) and let

$$P_i = a + b(M_i - 6) + \epsilon_e, \tag{17}$$

TABLE 1

Comparison with Brillinger and Preisler (1985)

Parameter*	Brillinger and Preisler (1985)	This Paper
$a - 6b$	−1.229	−1.229
b	0.277	0.277
c	−0.00231	−0.00231
h	6.650	6.650
σ_r†	0.2284	0.2283
σ_e†	0.1223	0.1222
$\hat{\sigma}_r$‡		0.2309
$\hat{\sigma}_e$‡		0.1236

* Parameter values correspond to the use of logarithms to the base 10 in equation (1).

† Maximum-likelihood estimate.

‡ Based on an unbiased estimate of σ^2.

where M_i is the magnitude of earthquake i. Linearizing the problem as before, we set

$$\mathbf{Y_1} = \begin{bmatrix} \log A_1 + \log\left(d_1^{\,2} + h'^2\right)^{1/2} \\ \log A_2 + \log\left(d_2^{\,2} + h'^2\right)^{1/2} \\ \vdots \\ \log A_N + \log\left(d_N^{\,2} + h'^2\right)^{1/2} \end{bmatrix}, \tag{18}$$

$$\mathbf{B_1} = \begin{bmatrix} c \\ \Delta h \\ P_1 \\ \vdots \\ P_{N_e} \end{bmatrix}, \tag{19}$$

and

$$\begin{bmatrix} (d_1^{\,2} + h'^2)^{1/2} & \left\{\frac{\partial}{\partial h}\left[c'(d_1^{\,2} + h^2)^{1/2} - \log(d_1^{\,2} + h^2)^{1/2}\right]\right\}_{h=h'} & E_{11} & E_{12} & \cdots & E_{1N_e} \\ (d_2^{\,2} + h'^2)^{1/2} & \left\{\frac{\partial}{\partial h}\left[c'(d_2^{\,2} + h^2)^{1/2} - \log(d_2^{\,2} + h^2)^{1/2}\right]\right\}_{h=h'} & E_{21} & E_{22} & \cdots & E_{2N_e} \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ (d_N^{\,2} + h'^2)^{1/2} & \left\{\frac{\partial}{\partial h}\left[c'(d_N^{\,2} + h^2)^{1/2} - \log(d_N^{\,2} + h^2)^{1/2}\right]\right\}_{h=h'} & E_{N1} & E_{N2} & \cdots & E_{NN_e} \end{bmatrix}, \tag{20}$$

where $E_{ni} = 1$ if recording n comes from earthquake i and $E_{ni} = 0$ otherwise. As before, N is the total number of data points and h' and c' are trial values of h and c. In each new iteration, $h' + \Delta h$ replaces h' of the previous iteration. The initial value of h' must be nonzero positive. Equation (1) can be replaced by the system

$$\mathbf{Y_1} = \mathbf{X_1}\mathbf{B_1} + \mathbf{e_1}, \tag{21}$$

where $\mathbf{e_1}$ is the vector of deviations with components ϵ_r. Since the components of $\mathbf{e_1}$ are assumed to be independent Gaussian random variables with zero mean and variance $\sigma_r^{\,2}$, the maximum-likelihood solution for $\mathbf{B_1}$ is the same as the ordinary least-squares solution (Searle, 1971, p. 87),

$$\hat{\mathbf{B}}_1 = \left(\mathbf{X_1}^T\mathbf{X_1}\right)^{-1}\mathbf{X_1}^T\mathbf{Y_1}. \tag{22}$$

The matrix $\hat{\mathbf{B}}_1$ includes the estimates $\hat{P}_i$ of the quantities P_i. We can use the $\hat{P}_i$ to determine a and b in equation (1), but to do so correctly we must take account of the variance-covariance matrix of the $\hat{P}_i$. We can write

$$\hat{P}_i = a + b(M_i - 6) + \left(\hat{P}_i - P_i\right) + \epsilon_e. \tag{23}$$

Equation (23) may seem like a trivial rearrangement of equation (17), but it is an essential step. Since we will use the estimates $\hat{P}_i$ to determine a and b, we

need $\hat{P}_i$ on the left-hand side of the equation. Equation (23) shows the two distinct sources of the variance and covariance of $\hat{P}_i$, the error of estimate $\hat{P}_i - P_i$, and the intrinsic variability ϵ_e of the estimated quantity. If we set

$$\mathbf{Y_2} = \begin{bmatrix} \hat{P}_1 \\ \hat{P}_2 \\ \vdots \\ \hat{P}_{N_e} \end{bmatrix}, \tag{24}$$

$$\mathbf{B_2} = \begin{bmatrix} a \\ b \end{bmatrix}, \tag{25}$$

and

$$\mathbf{X_2} = \begin{bmatrix} 1 & M_1 - 6 \\ 1 & M_2 - 6 \\ \vdots & \vdots \\ 1 & M_{N_e} - 6 \end{bmatrix}, \tag{26}$$

equation (23) can then be written

$$\mathbf{Y_2} = \mathbf{X_2}\mathbf{B_2} + \mathbf{e_2}, \tag{27}$$

where $\mathbf{e_2}$ is the vector of deviations, which are composed of $(\hat{P}_i - P_i)$ and ϵ_e. $\hat{P}_i$ is the least-squares estimate of P_i found in the first-stage regression, so $\hat{P}_i$ is an unbiased estimate of P_i with error $(\hat{P}_i - P_i)$ that is a linear combination of the values of ϵ_r, which is an independent random variable. The error of estimate $(\hat{P}_i - P_i)$ is thereby uncorrelated with ϵ_e, and the variance-covariance matrix of $\mathbf{e_2}$ is given by

$$\mathbf{V_2} = \mathrm{var}(\hat{\mathbf{P}} - \mathbf{P}) + \sigma_e^2\mathbf{I}, \tag{28}$$

where $\mathrm{var}(\hat{\mathbf{P}} - \mathbf{P})$ is the variance-covariance matrix of the vector whose components are $(\hat{P}_i - P_i)$, $\mathbf{I}$ is the identity matrix, and $\sigma_e^2\mathbf{I}$ is the variance-covariance matrix of the vector whose components are ϵ_e. Since $\hat{P}_i$ is the element $(\hat{B}_1)_{i+2}$ of the vector $\hat{\mathbf{B}}_1$, and since P_i is the mean of $\hat{P}_i$,

$$\left[\mathrm{var}(\hat{\mathbf{P}} - \mathbf{P})\right]_{ik} = \left[\mathrm{var}(\hat{\mathbf{B}}_1)\right]_{i+2,\,k+2} \tag{29}$$

and can be obtained from the matrix

$$\mathrm{var}(\hat{\mathbf{B}}_1) = (\mathbf{X}_1{}^T\mathbf{X}_1)^{-1}\sigma_r^2 \tag{30}$$

(Searle, 1971, pp. 89–90).

The likelihood of the observations is

$$L_2 = (2\pi)^{-N_e/2}|\mathbf{V_2}|^{-1/2}\exp\left[-\tfrac{1}{2}(\mathbf{Y_2} - \mathbf{X_2}\mathbf{B_2})^T\mathbf{V_2}^{-1}(\mathbf{Y_2} - \mathbf{X_2}\mathbf{B_2})\right]. \tag{31}$$

Maximizing L_2 with respect to $\mathbf{B_2}$ is the equivalent of the generalized least-squares problem with the weighting matrix $\mathbf{V_2}^{-1}$ (Searle, 1971, p. 87). The solution is

$$\hat{\mathbf{B}}_2 = \left(\mathbf{X_2}^T \mathbf{V_2}^{-1} \mathbf{X_2}\right)^{-1} \mathbf{X_2}^T \mathbf{V_2}^{-1} \mathbf{Y_2}, \tag{32}$$

the only difficulty being that σ_e and therefore $\mathbf{V_2}$ are not known. However, the weighted least-squares problem is the equivalent of an ordinary least-squares problem in a transformed variable (Draper and Smith, 1981, pp. 108–109). By considering the equivalent ordinary least-squares problem, we can show (Searle, 1971, p. 93) that

$$E\left[\left(\mathbf{Y_2} - \mathbf{X_2}\hat{\mathbf{B}}_2\right)^T \mathbf{V_2}^{-1} \left(\mathbf{Y_2} - \mathbf{X_2}\hat{\mathbf{B}}_2\right)\right] = N - 2, \tag{33}$$

where $E[\ ]$ denotes expectation and $N - 2$ is the number of degrees of freedom, 2 being the rank of matrix $\mathbf{X_2}$. The left-hand side of equation (33) is an implicit function of σ_e through $\hat{\mathbf{B}}_2$ and $\mathbf{V_2}$ (see equations 28 and 32). We solve equation (33) for σ_e numerically by the method of bisection (Press *et al.*, 1989, p. 246). That method prescribes the choice of trial values of σ_e. For each trial value of σ_e, we solve equations (32) and (28) for $\hat{\mathbf{B}}_2$ and $\mathbf{V_2}$ and substitute in equation (33). The process is iterated until σ_e is determined to sufficient precision.

Variations of the two-stage method all have solutions of the form of equation (32) but with different weighting matrices. In our early work, we used diagonal weighting matrices (Joyner and Boore, 1981, 1982). In the case of peak horizontal velocity and response spectral values, each earthquake was given equal weight, and, in the case of peak horizontal acceleration, earthquakes recorded at only one site were given zero weight and all other earthquakes were given equal weight. Later we suggested (but did not use) a diagonal weighting matrix with each earthquake having weight w_i given by

$$w_i = \left(\sigma_r^2 / R_i + \sigma_e^2\right)^{-1}, \tag{34}$$

where R_i is the number of recordings for earthquake i and σ_e is determined iteratively with zero as the starting value (Joyner and Boore, 1988). Fukushima and Tanaka (1990) also used a diagonal weighting matrix. They weighted each earthquake by the number of recordings. It was Masuda and Ohtake (1992) who pointed out the need to include off-diagonal terms in the weighting matrix. Their weighting matrix was the inverse of $\mathrm{var}(\hat{P}_i - P_i)$. According to the analysis above, their weighting would be correct if $\sigma_e^2 = 0$. We will show in the next section, however, that such is not the case. Restating our conclusion in other words, Masuda and Ohtake (1992) neglected the contribution of the intrinsic variance of the estimated quantity P_i.

To avoid misunderstanding, we should point out that, although Masuda and Ohtake's weighting in the second stage is the equivalent of assuming $\sigma_e = 0$, their method is not the equivalent of setting $\epsilon_e = 0$ in equation (1), because ϵ_e is included in the quantities P_i evaluated in the first-stage regression. Setting $\epsilon_e = 0$ in equation (1) leads to an ordinary least-squares problem, but, as explained earlier, that is not what Masada and Ohtake did.

Results of applying the different weighting schemes in two-stage methods are given in Table 2. The parameters h, c, and σ_r are the same in all cases because

TABLE 2
COMPARISON OF TWO-STAGE METHODS

Parameters*	1	2	3	4	5	6
a	0.415	0.478	0.389	0.427	0.499	0.463
b	0.290	0.249	0.310	0.291	0.270	0.248
$\hat{\sigma}_e$	0.201	0.134	0.275	0.202		

* Parameter values correspond to the use of logarithms to the base 10 in equation (1). Other parameters in equation (1) are determined in the first-stage regression and are the same in all cases; the values are $c = -0.00255$, $h = 7.31$, and $\hat{\sigma}_r = 0.223$

1. Weighting in the second stage by the inverse of the matrix in equation (28).
2. Weighting in the second stage as used by Joyner and Boore (1981) for peak horizontal acceleration.
3. Uniform weighting in the second stage.
4. Weighting in the second stage as suggested by Joyner and Boore (1988), and given by equation (34) of the present paper.
5. Weighting in the second stage as used by Fukushima and Tanaka (1990).
6. Weighting in the second stage as used by Masuda and Ohtake (1992).

they are determined in the first stage, which is the same for all. For the other parameters the differences are not large, indicating that weighting does not have a large effect. As pointed out by Fukushima and Tanaka (1992), the off-diagonal terms in the weighting matrix for the second-stage regression, though required logically, are small. For the method given in column (1) of Table 2, the maximum absolute value of an off-diagonal term is 0.19 times the minimum absolute value of a diagonal term. As pointed out by Masuda and Ohtake (1992), the off-diagonal terms arise because the estimates $\hat{P}_i$ are mutually correlated as a consequence of the fact that they were determined in the first-stage regression along with the parameters h and c. To the extent to which that correlation can be neglected, weighting according to equation (34) is correct. The close agreement between columns 1 and 4 indicates that the correlation can be neglected, but there is little reason to do so. The calculations that produced the result in column 1 took only 17 sec of CPU time on a VAXstation 3100.

COMPARISON OF ONE-STAGE AND TWO-STAGE METHODS

To obtain an objective comparison of the two methods, we employ Monte Carlo simulation (Press *et al.*, 1989, p. 529). For each method we start with a set of assumed values for the parameters of equation (1) obtained by applying the method in question to the data set. We then take the magnitude and distance values from the data set and use the assumed parameters in equation (1) with the aid of a pseudorandom-number generator to simulate a set of acceleration values, which we process by the method in question. We then compute the means and standard deviations of the output parameters and compare with the assumed input values. We also use the output parameters to compute predictions at $M = 6.5$ and 7.5 for $d = 0$ and 25 km. The results based on 100 simulations are shown in Table 3.

TABLE 3

MONTE CARLO COMPARISON OF ONE-STAGE AND TWO-STAGE METHODS FOR PEAK ACCELERATION

Parameter*	Assumed Value	One-stage Mean of Simulations	Standard Deviation of Simulations	Assumed Value	Two-Stage Mean of Simulations	Standard Deviation of Simulations
a	0.431	0.428	0.043	0.415	0.411	0.053
b	0.277	0.289	0.047	0.290	0.302	0.059
c	−0.00231	−0.00235	0.00042	−0.00255	−0.00257	0.00043
h	6.65	6.62	1.28	7.31	7.34	1.41
Parameter*	Assumed Value	Median of Simulations	16–84 Percentile of Simulations	Assumed Value	Median of Simulations	16–84 Percentile of Simulations
$\hat{\sigma}_r$	0.231	0.233	0.221–0.248	0.223	0.223	0.211–0.238
$\hat{\sigma}_e$	0.124	0.109	0.069–0.139	0.201	0.197	0.146–0.238
Magnitude and Distance	Log Acceleration Calculated from Assumed Values	Mean Log Acceleration Calculated from Output Parameters	Standard Deviation of Log Acceleration Calculated from Output Parameters	Log Acceleration Calculated from Assumed Values	Mean Log Acceleration Calculated from Output Parameters	Standard Deviation of Log Acceleration Calculated from Output Parameters
$M = 7.5$, $d = 0$ km	0.008	0.029	0.111	−0.033	−0.012	0.124
$M = 6.5$, $d = 0$ km	−0.269	−0.257	0.082	−0.323	−0.314	0.086
$M = 7.5$, $d = 25$ km	−0.626	−0.617	0.080	−0.632	−0.619	0.096
$M = 6.5$, $d = 25$ km	−0.903	−0.903	0.043	−0.922	−0.921	0.053

* Parameter values corresponds to the use of logarithms to the base 10 in equation (1).

Both methods performed well judging by Table 3. Both methods are essentially unbiased, as can be seen by noting that, since there were 100 simulations, the standard deviation of the mean from the simulations is one tenth of the standard deviation of the individual values from the simulations. The largest amount by which a mean value from the simulations differs from the assumed value is only slightly greater than two standard deviations of the mean. The stochastic uncertainty of parameter values from the two methods is comparable, as can be seen by comparing standard deviations of individual parameter values from the simulations.

Note particularly the predictions at zero distance for magnitudes 6.5 and 7.5. There are very few data points in this range of magnitude and distance. It is very encouraging to note that there is no bias in the predictions. Furthermore, the standard deviation of the values calculated from the output parameters, which represents the contribution to prediction error from stochastic uncertainty in the parameters, is small compared to the assumed value of the standard deviation of the residuals from the simulations, which is given by $\sigma = (\sigma_r^2 + \sigma_e^2)^{1/2}$, indicating that σ is a reasonable approximation of the total prediction error.

To see how well the two methods do on a much sparser data set, we repeat the Monte Carlo comparison, this time for peak horizontal velocity with the data set we used in 1982 (Joyner and Boore, 1982, 1988), which includes 65 records from 13 earthquakes. For velocity there is an additional parameter s, which is added to the right-hand side of equation (1) for data recorded at soil sites. The results for 100 simulations are given in Table 4. The units of velocity are cm/sec. The results, though not quite so good as in Table 3, are quite satisfactory for both methods. For predictions at zero distance, the contribution to prediction error from stochastic uncertainty in the parameters is larger than for the peak acceleration data set, but still smaller than the value of σ from the simulations, indicating that σ is not a gross underestimate of the total prediction error.

The Monte Carlo tests indicate that both methods give satisfactory results for the data sets used in the tests. Of particular interest is a comparison of the two methods on the data used by Fukushima and Tanaka (1990), to see if the one-stage method is as successful as the two-stage method in separating the magnitude dependence from the distance dependence and avoiding the error produced by ordinary least-squares analysis of that data. Fukushima and Tanaka used the equation

$$\log A = a + bM_J + c \log d_h, \tag{35}$$

where A is peak horizontal acceleration (cm/sec^2), M_J is magnitude defined by the Japan Meterorological Agency, and d_h is hypocentral distance (km); a, b, and c are parameters chosen to fit the data. Table 5 shows the comparison of the one-stage maximum-likelihood method, the two-stage method, and the ordinary least-squares method applied to Fukushima and Tanaka's (1990) data. Weighting in the second stage of the two-stage method is by the inverse of the matrix in equation (28). The one-stage maximum-likelihood method gives essentially the same results as the two-stage method, and the results from the ordinary least-squares method differ from those of the two-stage method in the same way as found by Fukushima and Tanaka (1990). The results in Table 5 for the two-stage method are virtually identical to the those given by Fukushima

TABLE 4

MONTE CARLO COMPARISON OF ONE-STAGE AND TWO-STAGE METHODS FOR PEAK VELOCITY

Parameter*	Assumed Value	One-stage Mean of Simulations	Standard Deviation of Simulations	Assumed Value	Two-Stage Mean of Simulations	Standard Deviation of Simulations
a	2.160	2.200	0.111	2.150	2.185	0.090
b	0.467	0.464	0.119	0.461	0.459	0.095
c	−0.00192	−0.00217	0.00165	−0.00256	−0.00277	0.00155
h	4.05	4.53	1.49	4.00	4.37	1.37
s	0.197	0.175	0.066	0.167	0.152	0.062
Parameter*	Assumed Value	Median of Simulations	16–84 Percentile of Simulations	Assumed Value	Median of Simulations	16–84 Percentile of Simulations
$\hat{\sigma}_r$	0.216	0.233	0.205–0.260	0.199	0.199	0.178–0.219
$\hat{\sigma}_e$	0.236	0.242	0.185–0.312	0.166	0.151	0.090–0.232
Magnitude and Distance	Log Velocity Calculated from Assumed Values	Mean Log Velocity Calculated from Output Parameters	Standard Deviation of Log Velocity Calculated from Output Parameters	Log Velocity Calculated from Assumed Values	Mean Log Velocity Calculated from Output Parameters	Standard Deviation of Log Velocity Calculated from Output Parameters
$M = 7.5$, $d = 0$ km	2.442	2.429	0.207	2.396	2.395	0.181
$M = 6.5$, $d = 0$ km	1.975	1.965	0.153	1.935	1.936	0.142
$M = 7.5$, $d = 25$ km	1.605	1.610	0.162	1.540	1.549	0.130
$M = 6.5$, $d = 25$ km	1.138	1.146	0.094	1.079	1.090	0.073

* Parameter values correspond to the use of logarithms to the base 10 in equation (1).

TABLE 5

COMPARISON BETWEEN THE ONE-STAGE MAXIMUM-LIKELIHOOD METHOD, THE TWO-STAGE METHOD, AND THE ORDINARY LEAST-SQUARES METHOD APPLIED TO FUKUSHIMA AND TANAKA'S DATA

Parameter*	One-Stage Maximum-Likelihood	Two-Stage	Ordinary Least Squares
a	2.20	2.20	2.22
b	0.44	0.44	0.26
c	−1.73	−1.76	−1.19
$\hat{\sigma}_r$	0.26	0.26	
$\hat{\sigma}_e$	0.19	0.21	
$\hat{\sigma}$	0.33	0.34	0.33

* Parameter values correspond to the use of logarithms to the base 10 in equation (35).

and Tanaka (1990), indicating that the differences in weighting have little effect.

Of the two methods, only the two-stage method can readily be used with the techniques described by Toro (1981) and McLaughlin (1991) for overcoming the bias due to instruments that do not trigger. As explained by McLaughlin (1991, p. 271), there are difficulties in applying those techniques to data with correlated errors.

ACKNOWLEDGMENTS

We are grateful to Yoshimitsu Fukushima for sending us his data and to D. J. Andrews and Leif Wennerberg for reviewing the manuscript and suggesting changes that lead to significant improvement in clarity. This work was partially supported by the Nuclear Regulatory Commission.

REFERENCES

Abrahamson, N. A. and R. R. Youngs (1992). A stable algorithm for regression analysis using the random effects model, *Bull. Seism. Soc. Am.* **82,** 505–510.

Brillinger, D. R. and H. K. Preisler (1984). An exploratory analysis of the Joyner-Boore attenuation data, *Bull. Seism. Soc. Am.* **74,** 1441–1450.

Brillinger, D. R. and H. K. Preisler (1985). Further analysis of the Joyner-Boore attenuation data, *Bull. Seism. Soc. Am.* **75,** 611–614.

Campbell, K. W. (1981). Near-source attenuation of peak horizontal acceleration, *Bull. Seism. Soc. Am.* **71,** 2039–2070.

Campbell, K. W. (1989). *Empirical prediction of near-source ground motion for the Diablo Canyon power plant site, San Luis Obispo County, California, U.S. Geol. Surv. Open-File Rept. 89-484,* 115 p.

Draper, N. R. and H. Smith (1981). *Applied Regression Analysis,* 2nd ed., Wiley, New York, 709 pp.

Fukushima, Y. and T. Tanaka (1990). A new attenuation relation for peak horizontal acceleration of strong earthquake ground motion in Japan, *Bull. Seism. Soc. Am.* **80,** 757–783.

Fukushima, Y. and T. Tanaka (1992). Reply to T. Masuda and M. Ohtake's "Comment on 'A new attenuation relation for peak horizontal acceleration of strong earthquake ground motion in Japan,'" *Bull. Seism. Soc. Am.* **82,** 523.

Hanks, T. C. and H. Kanamori (1979). A moment magnitude scale, *J. Geophys. Res.* **84,** 2348–2350.

Hildebrand, F. B. (1965). *Methods of Applied Mathematics,* 2nd ed., Prentice-Hall, Englewood Cliffs, New Jersey, 362 pp.

Joyner, W. B. and D. M. Boore (1981). Peak horizontal acceleration and velocity from strong-motion records including records from the 1979 Imperial Valley, California, earthquake, *Bull. Seism. Soc. Am.* **71,** 2011–2038.

Joyner, W. B. and D. M. Boore (1982). Prediction of earthquake response spectra, *Proc. 51st Ann. Convention Structural Eng. Assoc. of Cal.*, also *U.S. Geol. Surv. Open-File Rept. 82-977*, 16 pp.

Joyner, W. B. and D. M. Boore (1988). Measurement, characterization, and prediction of strong ground motion, in *Proc. Conf. on Earthquake Eng. and Soil Dynamics II*, GT Div/ASCE, Park City, Utah, 27–30 June 1988; 43–102.

Masuda, T. and M. Ohtake (1992). Comment on "A new attenuation relation for peak horizontal acceleration of strong earthquake ground motion in Japan" by Y. Fukushima and T. Tanaka, *Bull. Seism. Soc. Am.* **82,** 521–522.

McLaughlin, K. L. (1991). Maximum likelihood estimation of strong-motion attenuation relationships, *Spectra* **7,** 267–279.

Press, W. H., B. P. Flannery, S. A. Teukolsky, and W. T. Vetterling (1989). *Numerical Recipes*, Cambridge University Press, Cambridge, 702 pp.

Richter, C. F. (1935). An instrumental earthquake magnitude scale, *Bull. Seism. Soc. Am.* **25,** 1–32.

Richter, C. F. (1958). *Elementary Seismology*, W. H. Freeman, San Francisco, 768 pp.

Searle, S. R. (1971). *Linear Models*, Wiley, New York, 532 pp.

Toro, G. R. (1981). Biases in seismic ground motion prediction, Massachusetts Institute of Technology, Department of Civil Eng. Res. Rept. R81-22, Cambridge, Massachusetts, 133 pp.

Appendix A: Accounting for the Component of Variability Associated with the Site

The variability in ground-motion data can be separated into three components: ϵ_e, an independent random variable that takes on a specific value for each earthquake; ϵ_s, an independent random variable that takes on a specific value for each recording site; and ϵ_o, an independent random variable that takes on a specific value for each record. The mean of ϵ_e over the population of earthquakes is zero, and the variance is σ_e^2; ϵ_e represents the earthquake-to-earthquake component of variability. The mean of ϵ_s over the population of sites is zero, and the variance is σ_s^2; ϵ_s represents the site-to-site component of variability. The mean of ϵ_o over the populations records is zero, and the variance is σ_o^2; ϵ_o represents the variability remaining after contributions from site and source have been accounted for. Because there are few records in the data set corresponding to different earthquakes recorded at the same site, we took advantage in the text, as explained there, of the simplification provided by lumping $\epsilon_s + \epsilon_o = \epsilon_r$. Here in the Appendix we do the analysis for the separate ϵ_s and ϵ_o. In equation (1) we replace ϵ_r by $\epsilon_s + \epsilon_o$. The variance of ϵ_r is σ_r^2, and $\sigma_r^2 = \sigma_s^2 + \sigma_o^2$. The only thing that changes for the one-stage method is the variance-covariance matrix $\mathbf{V}$. The covariance of observations of different earthquakes at different sites is zero, the covariance of observations of the same earthquake at different sites is σ_e^2, and the variance of individual observations is $\sigma^2 = \sigma_o^2 + \sigma_s^2 + \sigma_e^2 = \sigma_r^2 + \sigma_e^2$, just as before. The covariance of observations of different earthquakes at the same site, however, is now σ_s^2, whereas before it was zero. As before, we define a normalized matrix $\mathbf{v}$ by the equation $\sigma^2\mathbf{v} = \mathbf{V}$ and define $\gamma_e = \sigma_e^2/\sigma^2$. (In the text γ_e was denoted simply by γ.) We also define $\gamma_s = \sigma_s^2/\sigma^2$. We start out with $\mathbf{v}$ from equations (10) and (11) and with $\mathbf{v}^{-1}$ from equations (B1), (B2), and (B3). For the elements of $\mathbf{v}$ corresponding to observations of different earthquakes at the same site, we substitute γ_s, and we calculate the corresponding changes in $\mathbf{v}^{-1}$ with the Sherman-Morrison formula (Press *et al.*, 1989, pp. 66–68). Otherwise the solution is the same as before, except that the logarithm of the likelihood in equation (14) must now be maximized numerically with respect to two variables γ_e and γ_s instead of one. To find the maximum we use the downhill simplex method (Press *et al.*, 1989, pp. 289–293). In using the downhill simplex method, it is advisable to employ a

transformation of variables. The variables γ_e and γ_s are constrained to be positive, and $\gamma_e + \gamma_s \leqq 1$ because $\sigma_e^2 + \sigma_s^2 + \sigma_o^2 = \sigma^2$. The subroutine given by Press *et al.*, (1989) for implementing the downhill simplex method, however, presumes that the independent variables may take on any real values. We use the following transformation:

$$p = \left(\frac{\gamma_e[\ln(1 + \gamma_e + \gamma_s) - \ln(1 - \gamma_e - \gamma_s)]}{2(\gamma_e + \gamma_s)} \right)^{1/2},$$

$$q = \left(\frac{\gamma_s[\ln(1 + \gamma_e + \gamma_s) - \ln(1 - \gamma_e - \gamma_s)]}{2(\gamma_e + \gamma_s)} \right)^{1/2}. \tag{A1}$$

The inverse transformation is

$$\gamma_e = \frac{p^2 \tanh(p^2 + q^2)}{(p^2 + q^2)},$$

$$\gamma_s = \frac{q^2 \tanh(p^2 + q^2)}{(p^2 + q^2)}. \tag{A2}$$

The results are compared in Table A1 with the results of the one-stage method described in the text. The differences are probably not significant.

For the analysis of a data set including both horizontal components, a strategy similar to that described above can be used to accommodate the covariance between orthogonal components recorded for the same earthquake at the same site.

In the two-stage method, the difference is that the variance-covariance matrix of the observations in the first stage is no longer diagonal, so the ordinary least-squares method may no longer be used. Since the components of $\mathbf{e}_1$ in equation (21) are normally distributed with zero mean and variance-covariance matrix of $\mathbf{V}_1$, the likelihood of the sample of observations is

$$L_1 = (2\pi)^{-N/2} |\mathbf{V}_1|^{-1/2} \exp\left[-\tfrac{1}{2} (\mathbf{Y}_1 - \mathbf{X}_1\mathbf{B}_1)^T \mathbf{V}_1^{-1} (\mathbf{Y}_1 - \mathbf{X}_1\mathbf{B}_1) \right] \tag{A3}$$

TABLE A1

COMPARISON OF RESULTS BETWEEN METHODS DESCRIBED IN THE TEXT AND IN APPENDIX A

Parameter*	One-Stage Appendix A	One-Stage Text	Two-Stage Appendix A	Two Stage Text
a	0.454	0.431	0.417	0.415
b	0.256	0.277	0.289	0.290
c	−0.00217	−0.00231	−0.00255	−0.00255
h	7.08	6.65	7.34	7.31
$\hat{\sigma}_o$	0.188		0.214	
$\hat{\sigma}_s$	0.148		0.063	
$\hat{\sigma}_r$	0.239	0.231	0.223	0.223
$\hat{\sigma}_e$	0.085	0.124	0.203	0.201

* Parameter values correspond to the use of logarithms to the base 10 in equation (1).

(Searle, 1971, p. 87). For a given $\mathbf{V_1}$, maximizing L_1 with respect to $\mathbf{B_1}$ yields

$$\hat{\mathbf{B}}_1 = \left(\mathbf{X_1}^T \mathbf{V_1}^{-1} \mathbf{X_1}\right)^{-1} \mathbf{X_1}^T \mathbf{V_1}^{-1} \mathbf{Y_1} \tag{A4}$$

(Searle 1971, p. 87). We define a normalized matrix $\mathbf{v_1}$ by the equation $\sigma_r^2 \mathbf{v_1} = \mathbf{V_1}$, where $\sigma_r^2 = \sigma_s^2 + \sigma_o^2$, and we define $\gamma_s = \sigma_s^2/\sigma_r^2$. The covariance of observations at different sites is zero, the covariance of observations of different earthquakes at the same site is σ_s^2, and the variance of an individual observation is σ_r^2. To determine the matrices $\mathbf{v_1}$ and $\mathbf{v_1}^{-1}$, we start with the identity matrix for both. For the elements of $\mathbf{v_1}$ corresponding to observations of different earthquakes at the same site, we substitute γ_s, and we calculate the corresponding changes in $\mathbf{v_1}^{-1}$ with the Sherman-Morrison formula (Press *et al.*, 1989, pp. 66–68). Substituting $\sigma_r^2 \mathbf{v_1}$ for $\mathbf{V_1}$ in equation (A4) gives

$$\hat{\mathbf{B}}_1 = \left(\mathbf{X_1}^T \mathbf{v_1}^{-1} \mathbf{X_1}\right)^{-1} \mathbf{X_1}^T \mathbf{v_1}^{-1} \mathbf{Y_1}. \tag{A5}$$

Substituting $\sigma_r^2 \mathbf{v_1}$ for $\mathbf{V_1}$ in equation (A3) and taking the natural logarithm gives

$$\ln L_1 = -\frac{N}{2}\ln(2\pi) - \frac{N}{2}\ln\left(\sigma_r^2\right) - \frac{1}{2}\ln|\mathbf{v_1}| - \frac{1}{2}(\mathbf{Y_1} - \mathbf{X_1}\mathbf{B_1})^T \mathbf{v_1}^{-1}(\mathbf{Y_1} - \mathbf{X_1}\mathbf{B_1})/\sigma_r^2. \tag{A6}$$

The likelihood L_1 must be maximized over all γ_s, $\mathbf{B_1}$, h, and σ_r^2. For fixed γ_s, L_1 can be maximized with respect to $\mathbf{B_1}$ and h by iterating on equation (A5) until $|\Delta h/h|$ is reduced below a specified limit, generally 10^{-3}. Since $\hat{\mathbf{B}}_1$ and h do not depend on σ_r^2 for fixed γ_s, we may then proceed to maximize L_1 with respect to σ_r^2 without recomputing $\hat{\mathbf{B}}_1$ or h. We differentiate equation (A6) with respect to σ_r^2, set the result equal to zero, and solve for σ_r^2. The solution is

$$\sigma_r^2 = \left(\mathbf{Y_1} - \mathbf{X_1}\hat{\mathbf{B}}_1\right)^T \mathbf{v_1}^{-1}\left(\mathbf{Y_1} - \mathbf{X_1}\hat{\mathbf{B}}_1\right)/N. \tag{A7}$$

For each value of γ_s, we compute values of $\hat{\mathbf{B}}_1$, h, σ_r^2, and the likelihood L_1 maximized with respect to $\hat{\mathbf{B}}_1$, h, and σ_r^2. The final solution corresponds to the value of γ_s for which the logarithm of the likelihood (equation A6) is maximum. The solution is found numerically by searching over γ_s using the search routine GOLDEN given by Press *et al.* (1989, p. 282).

The value of σ_r obtained from equation (A7) is not unbiased. An unbiased estimate is given by

$$\hat{\sigma}_r^2 = \left(\mathbf{Y_1} - \mathbf{X_1}\hat{\mathbf{B}}_1\right)^T \mathbf{v}^{-1}\left(\mathbf{Y_1} - \mathbf{X_1}\hat{\mathbf{B}}_1\right)/(N - N_e - 2), \tag{A8}$$

where $N - N_e - 2$ is the number of degrees of freedom. This result is obtained by noting that, for the final value of γ_s, the determination of $\hat{\mathbf{B}}_1$ is a weighted least-squares problem with weighting matrix $\mathbf{v}^{-1}$. Draper and Smith (1981, pp. 108–109) have shown that such a problem is the equivalent of an ordinary least-squares problem in a transformed variable. Equation (A8) follows from

applying the usual rules (Searle, 1971, p. 93) to the equivalent ordinary least-squares problem.

The second stage of the two-stage method is done in the same way as described in the text. The final results are compared in Table A1 with the results from the two-stage method described in the text.

As with the one-stage method, the two-stage method can be modified to accommodate a data set including both horizontal components.

Appendix B: Analytical Determination of the Inverse and the Determinant of the Normalized Variance-Covariance Matrix for the One-Stage Method

Since the rank of the matrix $\mathbf{v}$ is equal to the total number of recordings in the data set, it is advantageous to determine the inverse and determinant analytically. The inverse of $\mathbf{v}$ is given by

$$\mathbf{v}^{-1} = \begin{bmatrix} \mathbf{v_1}^{-1} & 0 & \cdots & 0 \\ 0 & \mathbf{v_2}^{-1} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \mathbf{v}_{\mathbf{N_e}}^{-1} \end{bmatrix}, \tag{B1}$$

as can be verified by carrying out multiplication by $\mathbf{v}$. The diagonal elements of $\mathbf{v_i}^{-1}$ are

$$\frac{1 + (R_i - 2)\gamma}{1 - (R_i - 1)\gamma^2 + (R_i - 2)\gamma}, \tag{B2}$$

and the off-diagonal elements

$$\frac{-\gamma}{1 - (R_i - 1)\gamma^2 + (R_i - 2)\gamma}, \tag{B3}$$

as can be verified by performing the multiplication $\mathbf{v_i}\mathbf{v_i}^{-1}$. The determinant $|\mathbf{v}|$ is equal to the product $|\mathbf{v_1}||\mathbf{v_2}|\ldots|\mathbf{v}_{\mathbf{N_e}}|$. That proposition follows directly from the definition of the determinant (e.g., Hildebrand, 1965, p. 10). Since the matrix $\mathbf{v_i}$ depends on i only through its rank R_i, we may change the notation to $\mathbf{v}_{\mathbf{R_i}}$. If $R_i = 1$, $|\mathbf{v}_{\mathbf{R_i}}| = 1$. If $R_i = 2$, $|\mathbf{v}_{\mathbf{R_i}}| = 1 - \gamma^2$. In general,

$$|\mathbf{v}_{\mathbf{R_i}}| = |\mathbf{v}_{\mathbf{R_i-1}}|\frac{1 - (R_i - 1)\gamma^2 + (R_i - 2)\gamma}{1 + (R_i - 2)\gamma}. \tag{B4}$$

To obtain this result, we note that a diagonal element of $\mathbf{v_i}^{-1}$, given by equation (B2), is also equal to

$$\frac{|\mathbf{v}_{\mathbf{R_i-1}}|}{|\mathbf{v}_{\mathbf{R_i}}|}, \tag{B5}$$

by the well-known equation (Hildebrand, 1965, pp. 16–17) giving the elements of an inverse matrix in terms of the cofactors of elements of the original matrix,

because $|\mathbf{v}_{\mathbf{R}_i-1}|$ is the cofactor of a diagonal element of $\mathbf{v}_i$. Equation (B4) can be rewritten

$$|\mathbf{v}_{\mathbf{R}_i}| = |\mathbf{v}_{\mathbf{R}_i-1}|(1-\gamma)\frac{1+(R_i-1)\gamma}{1+(R_i-2)\gamma}, \tag{B6}$$

which leads to

$$|\mathbf{v}_{\mathbf{R}_i}| = \prod_{n=2}^{R_i}(1-\gamma)\frac{1+(n-1)\gamma}{1+(n-2)\gamma} \tag{B7}$$

and, finally,

$$|\mathbf{v}_{\mathbf{R}_i}| = (1-\gamma)^{R_i-1}(1+[R_i-1]\gamma). \tag{B8}$$

Searle (1971, p. 462) gives expressions for $\mathbf{V}^{-1}$ and $|\mathbf{V}|$ that are equivalent to our equations (9), (B1), (B2), (B3), and (B8). Searle's expressions are used by Abrahamson and Youngs (1992). The essential difference in our approach is that we express things in terms of σ^2 and γ instead of σ_r^2 and σ_e^2 and, as a result, need to do numerical maximization of the likelihood over only one variable γ.

U. S. GEOLOGICAL SURVEY
345 MIDDLEFIELD ROAD
MS977
MENLO PARK, CALIFORNIA 94025

Manuscript received 8 May 1992