Ann. Inst. Statist. Math.
Vol. 45, No. 3, 531–540 (1993)

ESTIMATION OF A STRUCTURAL LINEAR REGRESSION MODEL WITH A KNOWN RELIABILITY RATIO*

HELENO BOLFARINE AND LISBETH K. CORDANI

Departamento de Estatistica, Universidade de São Paulo, Caixa Postal 20570, CEP 01452-990-SP, Brasil

(Received August 13, 1990; revised January 18, 1993)

Abstract. In this paper, we consider the estimation of the slope parameter β of a simple structural linear regression model when the reliability ratio (Fuller (1987), *Measurement Error Models*, Wiley, New York) is considered to be known. By making use of an orthogonal transformation of the unknown parameters, the maximum likelihood estimator of β and its asymptotic distribution are derived. Likelihood ratio statistics based on the profile and on the conditional profile likelihoods are proposed. An exact marginal posterior distribution of β, which is shown to be a t-distribution is obtained. Results of a small Monte Carlo study are also reported.

Key words and phrases: Orthogonality, profile likelihood, measurement error model, conditional model, likelihood ratio statistic, marginal posterior distribution.

1. Introduction

The classical structural simple linear regression model is defined by the equations

$$\begin{aligned} y_k &= \alpha + \beta x_k, \\ Y_k &= y_k + e_k, \\ X_k &= x_k + u_k \end{aligned} \tag{1.1}$$

where e_k, u_k and x_k are independent and normally distributed with means 0, 0, and μ_x, and variances σ_e^2, σ_u^2 and σ_x^2, respectively, that is, $e_k \sim N(0, \sigma_e^2)$, $u_k \sim N(0, \sigma_u^2)$ and $x_k \sim N(\mu_x, \sigma_x^2)$, $k = 1, \dots, n$. The main idea behind the equations (1.1) is that $(x_1, y_1), \dots, (x_n, y_n)$ are not observed directly and the estimation has to be based on $(X_1, Y_1), \dots, (X_n, Y_n)$. Extensive bibliographies on the structural and functional simple regression models are given in Kendall and Stuart (1961), Sprent (1990) and Fuller (1987). Examples of practical situations where the x_k are not

* The first author acknowledges partial finantial suport from CNPq-BRASIL.

observed directly are provided in Fuller (1987). An interesting situation is the case where x_k is the amount of nitrogen in the soil and y_k is the yield of a certain cereal. Values for X_k in this case can only be determined by laboratory analysis and are only estimates of the true x_k values.

From the previous assumptions, it follows that the joint distribution of (Y_k, X_k) is bivariate normal, that is,

$$\begin{pmatrix} Y_k \\ X_k \end{pmatrix} \sim N_2\left(\begin{pmatrix} \alpha+\beta\mu_x \\ \mu_x \end{pmatrix}; \begin{pmatrix} \beta^2\sigma_x^2+\sigma_e^2 & \beta\sigma_x^2 \\ \beta\sigma_x^2 & \sigma_x^2+\sigma_u^2 \end{pmatrix}\right), \tag{1.2}$$

$k = 1, \ldots, n$. Bayesian estimation of β when the ratio $\lambda = \sigma_e^2/\sigma_u^2$ is known and $\mu_x = 0$ is considered in Lindley and El-Sayad (1968). Likelihood based inferences are considered in Fuller (1987) and Wong (1989). In the present paper we assume that the reliability ratio,

$$k_x = \frac{\sigma_x^2}{\sigma_x^2+\sigma_u^2}, \tag{1.3}$$

is known. As pointed out by Fuller (1987), there is a great number of situations, particularly in psychology, sociology and survey sampling where k_x is so well estimated that it may be treated as being known. Table 1.1.1 in Fuller (1987) describes values of k_x for several variables. For example, measurement error is about 15% of the observed variation for income. By considering k_x as known, we may write $\sigma_u^2 = k\sigma_x^2$, where $k = (1-k_x)/k_x$. Hence, the covariance matrix in (1.2) may be written as

$$\begin{pmatrix} \beta^2\sigma_x^2+\sigma_e^2 & \beta\sigma_x^2 \\ \beta\sigma_x^2 & (k+1)\sigma_x^2 \end{pmatrix}. \tag{1.4}$$

In Section 2, we consider an orthogonal parametrization (Cox and Reid (1987)) of the unknown parameters which simplifies considerably the task of deriving the maximum likelihood estimators and approximate confidence intervals for β. Likelihood ratio statistics based on the profile and on the conditional profile likelihood (Cox and Reid (1987)) are also considered. Additionally, we perform a small Monte Carlo study in order to illustrate the performance of the confidence intervals based on the asymptotic distribution of the maximum likelihood estimator of β and also on the profile and conditional profile likelihoods. A conditional model which yields an exact confidence interval for β is proposed in Section 3. In Section 4 we consider the noninformative Jeffrey's priors for the unknown orthogonal parameters and show that the marginal posterior distribution of β is a t-distribution centered about the maximum likelihood estimator. Thus, exact inference for β can also be based on this posterior distribution.

2. The orthogonal transformation

It follows from (1.2) and (1.3) that the log of the likelihood function of the unknown parameters μ_x, α, β, σ_x^2 and σ_e^2 is proportional to

$$
\begin{aligned}
(2.1)\qquad -\frac{n}{2}\log(D) - \frac{1}{2D}\Bigg\{ &(\beta^2\sigma_x^2+\sigma_e^2)\sum_{i=1}^{n}(X_i-\mu_x)^2 \\
&- 2\beta\sigma_x^2\sum_{i=1}^{n}(X_i-\mu_x)(Y_i-\alpha-\beta\mu_x) \\
&+\sigma_x^2(k+1)\sum_{i=1}^{n}(Y_i-\alpha-\beta\mu_x)^2\Bigg\},
\end{aligned}
$$

where

$$D = k\beta^2\sigma_x^4 + (k+1)\sigma_x^2\sigma_e^2.$$

The main interest in the present paper is to make inferences about the slope parameter β, considering the other unknown parameters as nuisance parameters. As pointed out by Cox and Reid (1987), inference about β is typically simplified by an orthogonal parametrization, where β is orthogonal to the new parameters λ_0, λ_1, λ_2, λ_3. After computing the elements of the Fisher information matrix needed for finding the orthogonal parametrization $(\beta, \lambda_0, \lambda_1, \lambda_2, \lambda_3)$, we arrive at the following differential equations:

$$
\begin{aligned}
&[2k\beta^2 D + \sigma_e^4(k+1)^2]\frac{\partial\sigma_x^2}{\partial\beta} + \beta^2\sigma_x^4 k(k+1)\frac{\partial\sigma_e^2}{\partial\beta} = -2\beta^3\sigma_x^6 k^2,\\
&k(k+1)\beta^2\frac{\partial\sigma_x^2}{\partial\beta} + (k+1)^2\frac{\partial\sigma_e^2}{\partial\beta} = -2k(k+1)\sigma_x^2\beta,\\
&[\sigma_e^2 + k\beta^2\sigma_x^2]\frac{\partial\mu_x}{\partial\beta} + k\beta\sigma_x^2\frac{\partial\alpha}{\partial\beta} = -k\mu_x\beta\sigma_x^2
\end{aligned}
$$

and

$$k\beta\frac{\partial\mu_x}{\partial\beta} + (k+1)\frac{\partial\alpha}{\partial\beta} = -(k+1)\mu_x.$$

One solution to this set of differential equations gives the one to one transformation

$$
(2.2)\qquad \begin{cases}
\lambda_0 = \mu_x,\\
\lambda_1 = \alpha + \beta\mu_x,\\
\lambda_2 = (k+1)\sigma_e^2 + k\lambda_1\beta^2,\\
\lambda_3 = \sigma_x^2.
\end{cases}
$$

Therefore, it follows from (2.1) that the log of the likelihood function in the new parametrization is proportional to

$$
(2.3)\qquad -\frac{n}{2}\log(\lambda_2\lambda_3) - \frac{1}{2\lambda_2\lambda_3}\Bigg\{\frac{(\lambda_2+\lambda_3\beta^2)}{k+1}\sum_{i=1}^{n}(X_i-\lambda_0)^2
$$

$$- 2\beta\lambda_3 \sum_{i=1}^{n}(X_i - \lambda_0)(Y_i - \lambda_1)$$
$$\left. + \lambda_3(k+1)\sum_{i=1}^{n}(Y_i - \lambda_1)^2 \right\}.$$

After some algebraic manipulations, we arrive at the following maximum likelihood estimators:

$$\begin{aligned} \hat{\beta} &= (k+1)\frac{\sum_{i=1}^{n}(X_i - \bar{X})(Y_i - \bar{Y})}{\sum_{i=1}^{n}(X_i - \bar{X})^2} \\ &= k_x^{-1}\hat{\beta}_{OLS}, \\ \hat{\lambda}_0 &= \bar{X}, \\ \hat{\lambda}_1 &= \bar{Y}, \\ \hat{\lambda}_2 &= k_x\hat{\beta}^2 S_X^2 - 2\hat{\beta}S_{XY} + k_x^{-1}S_Y^2, \\ \hat{\lambda}_3 &= k_x S_X^2, \end{aligned} \tag{2.4}$$

where $\hat{\beta}_{OLS}$ denotes the ordinary least squares estimator of β, $S_X^2 = \sum_{i=1}^{n}(X_i - \bar{X})^2/n$, $S_Y^2 = \sum_{i=1}^{n}(Y_i - \bar{Y})^2/n$ and $S_{XY} = \sum_{i=1}^{n}(X_i - \bar{X})(Y_i - \bar{Y})/n$. Estimator (2.4) appears in Fuller (1987), and was suggested by arguments based on unbiasedness grounds. Using the above orthogonal parametrization we have shown that it is not only unbiased but it is also the maximum likelihood estimator of β. Thus, $\hat{\beta}$ is asymptotically efficient. Moreover, after some algebraic manipulations, it follows that the per observation Fisher information matrix in the new parametrization is given by

$$\begin{pmatrix} \frac{\beta^2\lambda_3 + \lambda_2}{(k+1)\lambda_2\lambda_3} & -\frac{\beta}{\lambda_2} & 0 & 0 & 0 \\ -\frac{\beta}{\lambda_2} & \frac{k+1}{\lambda_2} & 0 & 0 & 0 \\ 0 & 0 & \frac{1}{2\lambda_2^2} & 0 & 0 \\ 0 & 0 & 0 & \frac{1}{2\lambda_3^2} & 0 \\ 0 & 0 & 0 & 0 & \frac{\lambda_3}{\lambda_2} \end{pmatrix}. \tag{2.5}$$

Notice that, in this case, the orthogonal transformation (2.2) does not provide a global orthogonalization of the information matrix. An approximate $(1-\alpha)$ level confidence interval for β is given by

$$\left(\hat{\beta} - z_{\alpha/2}\sqrt{\frac{\hat{\lambda}_2}{n\hat{\lambda}_3}}; \hat{\beta} + z_{\alpha/2}\sqrt{\frac{\hat{\lambda}_2}{n\hat{\lambda}_3}}\right), \tag{2.6}$$

where $z_{\alpha/2}$ denotes the upper $\alpha/2$ point for the standard normal distribution. The above confidence interval is similar in form to the confidence interval (4) given in

Wong (1989). However, given the peculiarities of each model, we believe that one interval can not be obtained from the other. This point is also noted in Fuller (1987), where a separate analysis is considered for each model. Furthermore, it follows that the log of the profile likelihood, $l_p(\beta)$, for the parameter β is

$$l_p(\beta) \propto -\frac{n}{2}\log(\hat{\lambda}_2(\beta)), \tag{2.7}$$

where

$$\hat{\lambda}_2(\beta) = k_x \beta^2 S_X^2 - 2\beta S_{XY} + k_x^{-1} S_Y^2. \tag{2.8}$$

For testing the hypothesis $\beta = \beta_0$, we may use the likelihood ratio statistic, which is based on the profile likelihood above and is given by

$$W = 2\{l_p(\hat{\beta}) - l_p(\beta_0)\}, \tag{2.9}$$

where $l_p(\hat{\beta})$ and $l_p(\beta_0)$ are the profile likelihood (2.7) computed at the maximum likelihood estimator $\hat{\beta}$ and at β_0, respectively. Thus, a $(1-\alpha)$ level confidence interval for β can be obtained from $W < \chi_1^2(\alpha)$, where $\chi_1^2(\alpha)$ denotes the upper α level probability point for a chi-squared variate with one degree of freedom.

The log conditional profile likelihood function (Cox and Reid (1987)) of the parameter β, $l_c(\beta)$, is

$$l_c(\beta) \propto -\frac{(n-3)}{2}\log(\hat{\lambda}_2(\beta)), \tag{2.10}$$

where $\hat{\lambda}_2(\beta)$ is given in (2.8). Thus, we can obtain confidence intervals for β by using the conditional likelihood ratio statistic $W_c = 2\{l_c(\hat{\beta}) - l_c(\beta_0)\}$, as considered with the statistic W, given in (2.9).

Table 1 below presents the results of a Monte Carlo study based on 500 simulated samples generated according to model (1.1) with $\alpha = 1.0$, $\sigma_u^2 = 1.0$, $\beta = 1.0$ and 4.0, $k_x = 0.9$, 0.5 and 0.1, and $\mu_x = 0$ and 2.0. We took a significance level of 10% although simulation studies have indicated that there would be no significant difference in the results if other values such as 1% or 5% were taken. Table 1 shows that the approximate confidence interval I_1, given in (2.6), is generally conservative, that is, it presents a coverage level which is bigger than the nominal level in most of the cases considered in the simulation study. On the other hand, the coverage level presented by interval I_2 is, in most cases, below the nominal level. The best results seems to be achieved with the interval I_3, which presents coverage levels much closer to the nominal level than the other intervals. As expected, the confidence intervals based on the likelihood ratio statistics improve significantly for $n = 50$.

Table 1. Relative frequencies (x1000) where the confidence intervals (I_1, I_2 and I_3) covers $\beta = 1.0$ and 4.0.

	$\beta=1$						$\beta=4$					
	$k_x=0.9$		$k_x=0.5$		$k_x=0.1$		$k_x=0.9$		$k_x=0.5$		$k_x=0.1$	
	$\mu_x=0$	$\mu_x=2$	$\mu_x=0$	$\mu_x=2$	$\mu_x=0$	$\mu_x=2$	$\mu_x=0$	$\mu_x=2$	$\mu_x=0$	$\mu_x=2$	$\mu_x=0$	$\mu_x=2$
						$n=10$						
I_1	858	842	872	896	992	988	986	996	990	986	1000	998
I_2	854	816	824	830	840	826	844	850	868	836	840	850
I_3	918	904	908	906	908	916	902	908	934	900	898	910
						$n=50$						
I_1	954	970	974	980	1000	1000	1000	1000	1000	1000	1000	1000
I_2	898	902	880	880	874	874	890	898	866	912	904	882
I_3	908	908	886	902	882	852	900	908	876	916	916	896

(I_1: given by (2.6); I_2: based on (2.7); I_3 based on (2.10))

3. A conditional model

Considering the orthogonal transformation (2.2), it follows that the joint distribution of (Y_i, X_i) is bivariate normal with mean vector and covariance matrix given, respectively, by

$$\begin{pmatrix} \lambda_1 \\ \lambda_0 \end{pmatrix} \quad \text{and} \quad \begin{pmatrix} \dfrac{\lambda_2 + \lambda_3\beta^2}{k+1} & \beta\lambda_3 \\ \beta\lambda_3 & (k+1)\lambda_3 \end{pmatrix}.$$

Then, the conditional distribution of Y_i given X_i is

$$N(\alpha + \beta Z_i; k_x\lambda_2),$$

where $\alpha = \lambda_1 - \beta k_x \lambda_0$ and $Z_i = k_x X_i$, $i = 1, \ldots, n$.

Using standard least squares results, it follows that the estimators of β and λ_2 (given $\boldsymbol{X} = (X_1, \ldots, X_n)$), are given, respectively, by

$$\begin{aligned} \hat{\beta} &= \frac{\sum_{i=1}^n (Y_i - \bar{Y})(X_i - \bar{X})}{k_x \sum_{i=1}^n (X_i - \bar{X})^2} \quad \text{and} \\ \hat{\lambda}_2 &= \frac{1}{k_x(n-2)} \sum_{i=1}^n (Y_i - \bar{Y} - \hat{\beta} k_x (X_i - \bar{X}))^2. \end{aligned} \tag{3.1}$$

In this case, $\hat{\beta}$ is also a maximum likelihood estimator. However, the maximum likelihood estimator of λ_2 is as above, but corrected by $(n-2)n^{-1}$. Moreover, using standard results on the distribution of quadratic forms (see, for example, Searle (1971), p. 57), it can be shown that (conditional on $\boldsymbol{X}$) $(n-2)\hat{\lambda}_2/\lambda_2$ follows a chi-squared distribution with $n-2$ degrees of freedom. Since

$$\operatorname{Var}[\hat{\beta} \mid \boldsymbol{X}] = \frac{\lambda_2}{k_x \sum_{i=1}^n (X_i - \bar{X})^2},$$

which is estimated by replacing λ_2 by $\hat{\lambda}_2$ given in (3.1), it follows that a conditionally exact $(1-\alpha)$ level confidence interval for β is given by

$$\left(\hat{\beta} \pm t_{\alpha/2} \sqrt{\frac{\sum_{i=1}^n (Y_i - \bar{Y} - \hat{\beta} k_x (X_i - \bar{X}))^2}{(n-2) k_x^2 \sum_{i=1}^n (X_i - \bar{X})^2}} \right), \tag{3.2}$$

where $t_{\alpha/2}$ is the upper $\alpha/2$ point for the Student's t-distribution. This approch leads to the same results as found in Fuller ((1987), Section 1.1.2). This confidence interval extends a similar confidence interval derived by Rodrigues and Cordani (1990), under the less general assumption that μ_x and α are known.

4. A Bayesian analysis

As will be seen below, the orthogonal transformation (2.2) simplifies considerably the task of finding the exact marginal posterior distribution of the parameter β. We assign a uniform improper prior for $(\beta, \lambda_0, \lambda_1)$. Given $(\beta, \lambda_0, \lambda_1)$, we assign the Jeffrey's noninformative prior (Zellner (1971)) for (λ_2, λ_3). From the Fisher information matrix (2.5), the Jeffreys' noninformative prior for (λ_2, λ_3) is

$$p(\lambda_0, \lambda_1 \mid \beta) \propto \frac{1}{\lambda_2\lambda_3}.$$

Therefore, the joint posterior density of $(\beta, \lambda_0, \lambda_1, \lambda_2, \lambda_3)$ is proportional to

$$\begin{aligned}(4.1)\qquad & \left(\frac{1}{\lambda_2\lambda_3}\right)^{(n+2)/2} \exp\left\{-\frac{1}{2\lambda_3(k+1)}\sum_{i=1}^{n}(X_i-\lambda_0)^2\right\} \\ & \cdot \exp\left\{-\frac{1}{2\lambda_0}\left[\frac{\beta^2}{k+1}\sum_{i=1}^{2}(X_i-\lambda_0)^2\right.\right. \\ & \left.\left. -2\beta\sum_{i=1}^{n}(X_i-\lambda_0)(Y_i-\lambda_1)+(k+1)\sum_{i=1}^{n}(Y_i-\lambda_1)^2\right]\right\}.\end{aligned}$$

Integrating (4.1) over $(\lambda_0, \lambda_1, \lambda_2, \lambda_3)$, we arrive at the marginal posterior density of β, which is

$$\begin{aligned}(4.2)\qquad p(\beta \mid \boldsymbol{X}, \boldsymbol{Y}) \propto & \left\{\frac{\beta^2}{k+1}\sum_{i=1}^{n}(X_i-\bar{X})^2\right. \\ & -2\beta\sum_{i=1}^{n}(X_i-\bar{X})(Y_i-\bar{Y}) \\ & \left.+(k+1)\sum_{i=1}^{n}(Y_i-\bar{Y})^2\right\}^{-(n-1)/2},\end{aligned}$$

$\boldsymbol{Y} = (Y_1, \ldots, Y_n)'$. After some rearrangements, we can write the posterior density (4.2) as

$$(4.3)\qquad p(\beta \mid \boldsymbol{X}, \boldsymbol{Y}) \propto \left\{1+\frac{(\beta-\hat{\beta})^2}{(n-2)\Sigma}\right\}^{-(n-1)/2},$$

where

$$\Sigma = \frac{\{\sum_{i=1}^{n}(X_i-\bar{X})^2\sum_{i=1}^{n}(Y_i-\bar{Y})^2-(\sum_{i=1}^{n}(X_i-\bar{X})(Y_i-\bar{Y}))^2\}}{(n-2)k_x^2(\sum_{i=1}^{n}(X_i-\bar{X})^2)^2}.$$

Thus, according to (4.3), it follows that the posterior distribution of β is a Student's t-distribution with $n-2$ degrees of freedom and centered at the maximum likelihood estimator $\hat{\beta}$ given by (2.4). This means that given $\boldsymbol{X}$ and $\boldsymbol{Y}$, $t = (\beta - \hat{\beta})/\Sigma$

has a t-distibution with $n-2$ degrees of freedom so that a $100(1-\alpha)\%$ highest posterior density (HPD) interval for β is given by

$$(\hat{\beta}-t_{\alpha/2}\Sigma^{1/2};\hat{\beta}+t_{\alpha/2}\Sigma^{1/2}), \tag{4.4}$$

where $t_{\alpha/2}$ is the upper $\alpha/2$ point of a Student's t-distribution with $n-2$ degrees of freedom. It may be noted that the above (exact) HPD interval and the exact conditional interval (3.2) coincide. Moreover, the posterior variance of β is given by

$$\mathrm{Var}[\beta \mid \boldsymbol{X}, \boldsymbol{Y}] = \frac{(n-2)}{(n-4)}\Sigma. \tag{4.5}$$

As pointed out by Sweeting (1987), the posterior density of β given in (4.4) is close, for large sample sizes, to the conditional profile likelihood derived in the previous section.

Bayesian inference for β may also be obtained by using the conditional model of Section 3. By considering the noninformative prior

$$p(\beta,\lambda_2) \propto \frac{1}{\lambda_2},$$

it can be shown that the posterior marginal distribution of β is again a t-distribution with $n-2$ degrees of freedom and centered at $\hat{\beta}$ given in (2.4). The posterior variance is as given by (4.5) above.

Example. The 12 pairs (Y_i, X_i) presented next (Fuller (1987), Problem 5, Chapter 1), are supposed to follow model (1.1) with $k_x = 0.85$: $(3.6, 3.9)$, $(2.5, 2.9)$, $(3.9, 4.4)$, $(5.0, 5.9)$, $(4.9, 5.4)$, $(4.5, 4.2)$, $(2.9, 2.3)$, $(5.2, 4.5)$, $(2.7, 3.5)$, $(5.8, 6.0)$, $(4.1, 3.3)$, $(5.1, 4.1)$. After centering X_i and Y_i it follows that $\hat{\beta} = 0.9175$, $\hat{\lambda}_2 = 0.4038$ and $\hat{\lambda}_3 = 1.0200$ are the maximum likelihood estimators of β, λ_2 and λ_3. Now, by using (2.6), it follows that an approximate 95% confidence interval for β is given by $(0.5544; 1.2806)$. On the other hand, after some numerical manipulations, it follows that the HPD (4.4) (which coincides with the interval (3.2)) reduces to $(0.4741, 1.3609)$.

Acknowledgements

The authors wish to thank the referees for valuable comments and suggestions.

References

Cox, D. R. and Reid, N. (1987). Parameter orthogonality and approximate conditional inference (with discussion), *J. Roy. Statist. Soc. Ser. B*, **49**, 1–39.

Fuller, W. A. (1987). *Measurement Error Models*, Wiley, New York.

Kendall, M. G. and Stuart, A. (1961). *The Advanced Theory of Statistics*, Vol. 2, Griffin, London.

Lindley, D. V. and El Sayad, G. (1968). The Bayesian estimation of a linear functional relationship, *J. Roy. Statist. Soc. Ser. B*, **30**, 190–202.

Rodrigues, J. and Cordani, L. K. (1990). Conditional likelihood inference of a simple regression model with measurement error via the orthogonal reparametrization, *South African Statist. J.*, **24**, 177–183.

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

Sprent, P. (1990). Some history of functional and structural relationships, *Contemp. Math.*, **112**, 3–15.

Sweeting, T. J. (1987). Discussion of paper by D. R. Cox and N. Reid, *J. Roy. Statist. Soc. Ser. B*, **49**, 1–39.

Wong, M. Y. (1989). Likelihood estimation of a simple regression model when both variables have error, *Biometrika*, **76**, 141–148.

Zellner, A. (1971). *An Introduction to Bayesian Inference in Econometrics*, Wiley, New York.