The Journal of National Chengchi University Vol. 32, 1975

Bayesian Analysis of Linear Regression Model with Errors in the Variables

Ing-Tzer Wey

I. Introduction

Consider n pairs of observations (x_i, y_i), $i=1, 2, \ldots, n$, which are linearly related and both the variables are subject to error. The model for the observations can be written as

$$\left.\begin{aligned} x_i &= \mu_i + d_i \\ y_i &= \alpha + \beta\mu_i + e_i \end{aligned}\right\} = i = 1, 2, \ldots, n \tag{1}$$

where the errors d_i's and e_i's are independently and normally distributed with

$$\left.\begin{aligned} &E(d_i) = 0, \quad E(e_i) = 0 \\ &E(d_i^2) = \sigma_d^2, \; E(e_i^2) = \sigma_e^2 \\ \text{and} \quad &E(d_i e_i) = 0 \end{aligned}\right\} i = 1, 2, \ldots, n$$

α, β, μ_i's, σ_d^2 and σ_e^2 are unknown parameters. In the language of Neyman and Scott (1951) the parameters α, β, σ_d^2 and σ_e^2 are "structural", whereas the μ_i's are "incedental"; the distinction being that the structural parameters occur in the joint distribution of every observation pair, whereas μ_i is incedental to the pair (x_i, y_i) alone.

We shall first discuss the maximum likelihood approach to the estimation of parameters in the model (1) in order to compare it with the Bayesian approach.

II. The Maximum Likelihood Approach

Define the admissible parameter space for the model in (1) as follows: $0<\sigma_d^2, \sigma_e^2 <\infty$, $\sigma_d^2 \neq \sigma_e^2$; $-\infty<\alpha, \beta, \mu_i\text{'s}<+\infty$; $\beta^2 \neq \sigma_e^2/\sigma_d^2$. The likelihood function for the variables in the model (1) is given by

$$L(\underset{\sim}{x}, \underset{\sim}{y}:\alpha, \beta, \underset{\sim}{\mu}, \sigma_d, \sigma_e) \propto \frac{1}{\sigma_d^n \sigma_e^n} \exp\left[-\frac{1}{2\sigma_d^2} \sum_{i=1}^{n} (x_i - \mu_i)^2 - \frac{1}{2\sigma_e^2} \sum_{i=1}^{n} (y_i - \alpha - \beta\mu_i)^2 \right]$$

As a necessary condition for a maximum of the likelihood function, values of the parameters must exist in the admissible parameter space, which sets the partial derivatives of the likelihood function with respect to each of the parameters in turn equal to zero, and obtain the following results :

$$\hat{\underset{\sim}{\mu}} = \frac{1}{1+\hat{\theta}} (\underset{\sim}{x} + \hat{\theta}\, \underset{\sim}{w}) \tag{2}$$

where $\hat{\theta} = \hat{\beta}^2 \hat{\sigma}_d^2 / \hat{\sigma}_e^2$, $\hat{\underset{\sim}{\mu}} = (\hat{\mu}_1, \ldots, \hat{\mu}_n)'$, $\underset{\sim}{w} = (\underset{\sim}{y} - \hat{\alpha}\underset{\sim}{1}) / \hat{\beta}$, $\underset{\sim}{x} = (x_1, \ldots, x_n)'$, $\underset{\sim}{y} = (y_1, \ldots, y_n)'$ and $\underset{\sim}{1} = (1, \ldots, 1)'$ are n x 1 column vectors.

$$\hat{\sigma}_d^2 = \frac{\hat{\theta}^2}{n(1+\hat{\theta})^2} (\underset{\sim}{x} - \underset{\sim}{w})' (\underset{\sim}{x} - \underset{\sim}{w}) \tag{3}$$

$$\hat{\sigma}_e^2 = \frac{s^2}{n(1+\hat{\theta})^2} (\underset{\sim}{x} - \underset{\sim}{w})' (\underset{\sim}{x} - \underset{\sim}{w}) \tag{4}$$

These two equations can hold simultaneously if, and only if

$$\hat{\beta}^2 = \hat{\sigma}_e^2 / \hat{\sigma}_d^2 \tag{5}$$

However, it was explicitly stated in the definition of the admissible parameter space that $\beta^2 \neq \sigma_e^2 / \sigma_d^2$ and thus the quantity $\hat{\beta}^2 = \hat{\sigma}_e^2 / \hat{\sigma}_d^2$ falls in an inadmissible region of the parameter space. Hence, a maximum of the likelihood function does not exist in the admissible region of the parameter space. As noted by Kendall and Stuart (1961), Stein (1956), James and Stein (1961), and Stein (1962), this is obviously an absurd result.

Since there are basic difficulties with the analysis of the model in (1) when all parameters are assumed unknown, analysis has often gone forward under the assumption that $R = \sigma_e^2 / \sigma_d^2$ is known exactly. Under this assumption a unique maximum of the likelihood function exists and thus the maximum likelihood estimators have been obtained by Madansky (1959) as follows:

$$\hat{\beta} = \frac{1}{2m_{xy}} \left[m_{yy} - Rm_{xx} + \sqrt{(m_{yy} - Rm_{xx})^2 + 4Rm_{xy}^2} \right] \tag{6}$$

where $m_{xx} = \frac{1}{n} \sum_{i=1}^{n} (x_i - \bar{x})^2$, $m_{yy} = \frac{1}{n} \sum_{i=1}^{n} (y_i - \bar{y})^2$

$$m_{xy} = \frac{1}{n} \sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y})$$

$$\hat{\alpha} = \bar{y} - \hat{\beta}\bar{x} \tag{7}$$

$$\hat{\underset{\sim}{\mu}} = \frac{1}{R + \hat{\beta}^2} \left[R\underset{\sim}{x} + \hat{\beta}(\underset{\sim}{y} - \hat{\alpha}\underset{\sim}{1}) \right] \tag{8}$$

$$\hat{\sigma}_e^{\,2} = \frac{1}{2n} \left[R(\underset{\sim}{x} - \hat{\underset{\sim}{\mu}})'(\underset{\sim}{x} - \hat{\underset{\sim}{\mu}}) + (\underset{\sim}{y} - \hat{\alpha}\underset{\sim}{1} - \hat{\beta}\hat{\underset{\sim}{\mu}})'(\underset{\sim}{y} - \hat{\alpha}\underset{\sim}{1} - \hat{\beta}\hat{\underset{\sim}{\mu}}) \right] \tag{9}$$

It is easily verified that although α and β are consistent estimators, $\hat{\sigma}_e^{\,2}$ is not. In fact, $\text{plim}(\hat{\sigma}_e^{\,2}) = \frac{1}{2}\sigma_e^{\,2}$. The correction for inconsistency of the estimator $\hat{\sigma}_e^{\,2}$ can be made under the consideration that there are 2n observations and in (9) we have inserted estimates for n elements of μ, α and β, that is, for n+ 2 parmeters, then $2n - (n + 2) = n - 2$ represents the degrees of freedom remaining for the estimation of $\sigma_e^{\,2}$. Thus, a consistent estimator is $2n\,\hat{\sigma}_e^{\,2}/(n-2)$.

Now, consider further the model in (1) in which the μ_i's are assumed to be independent of the d_i's and the e_i's, and normally and independently distributed, each with mean τ and variance σ^2. We note that under this assumption along with the other distributional assumptions made in connection with (1), the pairs of variables (x_i, y_i), $i = 1, 2, \ldots, n$, are independently and identically distributed, each pair having a bivariate normal distribution with means and variance-covariances as follows:

$$\left.\begin{array}{ll} E(x_i) = \tau & E(y_i) = \alpha + \beta\tau \\ \text{Var}(x_i) = \sigma^2 + \sigma_d^{\,2}, & \text{Var}(y_i) = \beta^2\sigma^2 + \sigma_e^{\,2} \\ \text{Cov}(x_i, y_i) = \beta\sigma^2 & \end{array}\right\} \tag{10}$$

From these moments, with $\mathrm{Cov}(x_i, y_i) > 0$, we have

$$\frac{\mathrm{Cov}(x_i, y_i)}{\mathrm{Var}(x_i)} < \beta < \frac{\mathrm{Var}(y_i)}{\mathrm{Cov}(x_i, y_i)} \tag{11}$$

If $\mathrm{Cov}(x_i, y_i) < 0$, the inequalities in (11) are reversed. Thus the admissible values of β fall in one of two finite intervals, given by (11), or (11) with the inequalities reversed.

Since the five moments in (10)completely determine a bivariate normal distribution and the corresponding sample moments are sufficient statistics in this instance, we can equate sample moments to their respective population moments in an effort to obtain maximum likelihood estimates. There is, however, a basic difficulty with this approach, namely, that although we have five relations in (10), there are six unknown parameters τ, α, β, σ^2, σ_d^2 and σ_e^2. Thus we cannot obtain estimates of all parameters unless prior information is available to reduce the number of unknown parameters.

Assume first the case in which we know that $\alpha = 0$. When this information is available,we can quate sample moments to their respective population moments to obtain maximum likelihood estimates as follows:

$$\begin{aligned} \hat{\beta} &= \bar{y}/\bar{x} \\ \hat{\sigma}^2 &= m_{xy}/\hat{\beta} \\ \hat{\sigma}_e^2 &= m_{yy} - \hat{\beta}^2\hat{\sigma}^2 \\ \hat{\sigma}_d^2 &= m_{xx} - \hat{\sigma}^2 \end{aligned} \tag{12}$$

Although the information $\alpha = 0$ enables us to obtain estimates of the remaining parameters, it must be noted that this approach can lead to negative estimates for any or all of the following variances: σ_d^2, σ_e^2, and σ^2; that is, the prior information that these variances are positive has not been introduced explicitly and thus meaningless variance estimates can be obtained. Further, the estimate $\hat{\beta} = \bar{y} / \bar{x}$ for β has a range $-\infty$ to $+\infty$ and thus can violate the admissible bounds set by (11).

If, rather than α, $R = \sigma_e^2/\sigma_d^2$ is known, then maximum likelihood estimates can be obtained by equating sample moments to their respective population moments. Let $\mu_{11} = \mathrm{Var}(x_i)$, $\mu_{22} = \mathrm{Var}(y_i)$, and $\mu_{12} = \mathrm{Cov}(x_i, y_i)$, then

$$\sigma^2 = \mu_{12}/\beta$$

$$\sigma_d^2 = \mu_{11} - \sigma^2 = \mu_{11} - \mu_{12}/\beta$$

and $$\beta^2\mu_{12} + \beta(R\mu_{11} - \mu_{22}) - R\mu_{12} = 0$$

On replacing the μ_{ij}'s by their sample moments, we have

$$\beta^2 m_{xy} + \beta(Rm_{xx} - m_{yy}) - Rm_{xy} = 0$$

Thus the maximum likelihood estimator for β is a solution of this quadratic equation, namely

$$\hat{\beta} = \frac{1}{2m_{xy}}\left[m_{yy} - Rm_{xx} + \sqrt{(m_{yy} - Rm_{xx})^2 + 4Rm_{xy}^2}\right]$$

which is in precisely the same form as (6). Estimates for the remaining parameters are given by

$$\left.\begin{aligned} \hat{\sigma}^2 &= m_{xy}/\hat{\beta} \\ \hat{\sigma}_d^2 &= m_{xx} - \hat{\sigma}^2 \\ \hat{\sigma}_e^2 &= R\hat{\sigma}_d^2 \\ \hat{\alpha} &= \bar{y} - \hat{\beta}\bar{x} \\ \hat{\tau} &= \bar{x} \end{aligned}\right\} \qquad (13)$$

Above result is attained under the assumption that the μ_i's are normally, identically and independently distributed with mean τ and variance σ^2. With this assumption, additional prior information must be added to identify the parameters. However, if the μ_i's have a non-normal distribution, the parameters will be identified. To illustrate one case of non-normality assume that the probability density function(pdf) of μ_i is

$$p(\mu_i) \propto \text{const.} \quad i = 1, 2, \ldots, n \qquad (14)$$

In (14) we assume that the μ_i's are uniformly and independently distributed. Below we shall see how this assumption about the μ_i's affect maximum likelihood estimates.

The likelihood function of $\underset{\sim}{x}$, $\underset{\sim}{y}$, and $\underset{\sim}{\mu}$ is given by

$$L(\underset{\sim}{x}, \underset{\sim}{y}, \underset{\sim}{\mu} : \underset{\sim}{\theta}) \propto \frac{1}{\sigma_d^n \sigma_e^n} \exp\left[-\frac{1}{2\sigma_d^2} \sum_{i=1}^{n} (x_i - \mu_i)^2 - \frac{1}{2\sigma_e^2} \sum_{i=1}^{n} (y_i - \alpha - \beta\mu_i)^2 \right]$$

$$\propto \frac{1}{\sigma_d^n \sigma_e^n} \exp\left[-\frac{1}{2\sigma^2} \left\{ R \sum_{i=1}^{n} (x_i - \mu_i)^2 + \sum_{i=1}^{n} (y_i - \alpha - \beta\mu_i)^2 \right\} \right]$$

where $\underset{\sim}{\theta} = (\alpha, \beta, \sigma_d, R)'$. To perform the integration with respect to the elements of $\underset{\sim}{\mu}$ we can complete the square in the exponent and use properties of the normal distribution to obtain the marginal distribution of $\underset{\sim}{x}$ and $\underset{\sim}{y}$ as follows:

$$h(\underset{\sim}{x}, \underset{\sim}{y} : \underset{\sim}{\theta}) \propto \left[\frac{R}{\sigma_e^2 (R + \beta^2)} \right]^{\frac{n}{2}} \exp\left[- \frac{R}{2\sigma_e^2 (R + \beta^2)} \sum_{i=1}^{n} (y_i - \alpha - \beta\mu_i)^2 \right] \quad (15)$$

On maximizing ln h with respect to σ_e, α, and β, we find the maximum likelihood estimates for α and β to be

$$\left. \begin{aligned} \hat{\alpha} &= \bar{y} - \beta \bar{x} \\ \hat{\beta} &= \sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y}) \Big/ \sum_{i=1}^{n} (x_i - \bar{x})^2 \end{aligned} \right\} \quad (16)$$

It is seen that $\hat{\alpha}$ and $\hat{\beta}$ are just the simple least squares estimates obtaining from regressing $\underset{\sim}{y}$ on $\underset{\sim}{x}$. This result is vitally dependent on the assumption about the pdf for $\underset{\sim}{\mu}$ in (14), an assumption which implies that the μ_i's have infinite variances. Note that, in general, the probability limit of β given by (16) is as follows:

$$\text{plim}\ (\hat{\beta}) = \beta\ (1 + \sigma_d^2 / \sigma^2)^{-1}$$

Hence, as $\sigma^2 \rightarrow \infty$, plim $(\hat{\beta}) \rightarrow \beta$. Therefore $\hat{\beta}$ is a consistent estimator.

Ⅲ. The Bayesian Approach

One advantage of the Bayesian approach is that prior knowledge about parameters of interest can be combined in a well-defined mathematical way with information obtained from experiments. Such prior knowledge may arise from general theoretical considerations and/or the results of previous or concurrent experiments and is usually an important component of an investigator's quest for understanding. In this paper we illustrate how prior knowledge can be utilized in conjunction with sample information in making inferences about the parameters of the model in (1). As will be seen, there is no need to assume, for example, that the value of $R = \sigma_e^2/\sigma_d^2$ is known exactly in order to make inferences about the parameters of interest.

Let us first assume a priori that the μ_i's, α, β, $\ln\sigma_d$ and $\ln\sigma_e$ are uniformly and independently distributed with the following prior pdf:

$$p(\underset{\sim}{\mu}, \alpha, \beta, \sigma_d, \sigma_e) \propto \frac{1}{\sigma_d\sigma_e} \tag{17}$$

The likelihood function for parameters of the model in (1) is given by

$$L(\underset{\sim}{\mu}, \alpha, \beta, \sigma_d, \sigma_e \mid \underset{\sim}{x}, \underset{\sim}{y}) \propto \frac{1}{\sigma_d^n\sigma_e^n} \exp\left[-\frac{1}{2\sigma_d^2}\sum_{i=1}^n (x_i - \mu_i)^2 - \frac{1}{2\sigma_e^2}\sum_{i=1}^n (y_i - \alpha - \beta\mu_i)^2\right] \tag{18}$$

On combining the prior pdf in (17) with the likelihood function in (18), we have the following posterior pdf:

$$p(\underset{\sim}{\mu}, \alpha, \beta, \sigma_d, \sigma_e \mid \underset{\sim}{x}, \underset{\sim}{y}) \propto$$

$$\frac{1}{\sigma_d^{n+1}\sigma_e^{n+1}} \exp\left[-\frac{1}{2\sigma_d^2}\sum_{i=1}^n (x_i - \mu_i)^2 - \frac{1}{2\sigma_e^2}\sum_{i=1}^n (y_i - \alpha - \beta\mu_i)^2\right]$$

$$\propto \frac{1}{\sigma_d^{n+1}\sigma_e^{n+1}} \exp\left\{-\frac{1}{2\sigma_e^2}\left[R\sum_{i=1}^n (x_i - \mu_i)^2 + \sum_{i=1}^n (y_i - \alpha - \beta\mu_i)^2\right]\right\} \tag{19}$$

where $R = \sigma_e^2/\sigma_d^2$.

Completing the square on $\underset{\sim}{\mu}$ in the exponent of (19) and integrating with respect to the elements of $\underset{\sim}{\mu}$, we have

$$p(\alpha, \beta, \sigma_d, \sigma_e \mid \underset{\sim}{x}, \underset{\sim}{y}) \propto \frac{1}{\sigma_d^{n+1}\sigma_e(\beta^2 + R)^{n/2}} \times \exp\left[-\frac{1}{2\sigma_d^2(\beta^2 + R)}\sum_{i=1}^n (y_i - \alpha - \beta x_i)^2\right]$$

Eliminating the element σ_e from the above pdf, we have

$$p(\alpha, \beta, \sigma_d, R \mid \underset{\sim}{x}, \underset{\sim}{y}) \propto \frac{1}{R\sigma_d^{n+1}(\beta^2 + R)^{n/2}} \exp\left[-\frac{1}{2\sigma_d^2(\beta^2 + R)}\sum_{i=1}^n (y_i - \alpha - \beta x_i)^2\right]$$

Then, on integrating with respect to σ_d, we have

$$p(\alpha, \beta, R \mid \underset{\sim}{x}, \underset{\sim}{y}) \propto \frac{1}{R}\left[\sum_{i=1}^{n} (y_i - \alpha - \beta x_i)^2\right]$$

$$\propto \frac{1}{R}\left[vs^2 + (\hat{\underset{\sim}{\beta}} - \underset{\sim}{\beta})' X'X (\hat{\underset{\sim}{\beta}} - \underset{\sim}{\beta})\right]^{-\frac{n}{2}} \tag{20}$$

where $v = n-2$, $\underset{\sim}{\beta} = (\alpha, \beta)'$, $X = (\underset{\sim}{1}, \underset{\sim}{x})$, $\hat{\underset{\sim}{\beta}} = (X'X)^{-1} X'\underset{\sim}{y}$, and

$$vs^2 = (\underset{\sim}{y} - X\hat{\underset{\sim}{\beta}})' (\underset{\sim}{y} - X\hat{\underset{\sim}{\beta}})$$

From the form of (20) we see that $\underset{\sim}{\beta} = (\alpha, \beta)'$ has a posterior pdf in the bivariate Student t form with mean $\hat{\underset{\sim}{\beta}}$, precisely the least squares quantity obtained from regressing $\underset{\sim}{y}$ on $\underset{\sim}{x}$. As with the maximum likelihood approach yielding (16), the present result is critically dependent on the assumption about the form of the pdf for $\underset{\sim}{\mu}$ in (17). In addition, note that the posterior pdf for R in (20) is improper and exactly in the form implied by the prior assumptions about σ_d and σ_e in (17). Thus, with the prior assumptions in (17), no new information is provided about R from the sample.

Although the assumptions embodied in (17) permit us to make posterior inferences about α and β which are appropriate in certain situations, they, of course, are not appropriate in all problems. In certain circumstances, we may find it appropriate to assume that the n elements of $\underset{\sim}{\mu}$ are, a prior, normally and independently distributed with a common mean τ and variance σ^2; that is, the pdf of $\underset{\sim}{\mu}$ is given by

$$p(\underset{\sim}{\mu} \mid \tau, \sigma^2) \propto \frac{1}{\sigma^n} \exp\left[-\frac{1}{2\sigma^2} \sum_{i=1}^{n} (\mu_i - \tau)^2\right] \tag{21}$$

If we assume further that τ and σ are to be unknown and distributed a priori as $p(\tau, \sigma) \propto$ const, with $-\infty < \tau < +\infty$ and $0 < \sigma < +\infty$, then the marginal prior pdf for $\underset{\sim}{\mu}$ is

$$p(\underset{\sim}{\mu}) \propto \left[\sum_{i=1}^{n} (\mu_i - \bar{\mu})^2\right]^{-(n-2)/2} \qquad -\infty < \mu_i < \infty \tag{22}$$

$$i = 1, 2, \ldots, n$$

where $\quad \bar{\mu} = \frac{1}{n} \sum_{i=1}^{n} \mu_i .$

The likelihood function is

$$L(\alpha, \beta, R, \sigma_e, \underset{\sim}{\mu} \mid \underset{\sim}{x}, \underset{\sim}{y}) \propto$$

$$\frac{R^{n/2}}{\sigma_e^{2n}} \exp\left\{-\frac{1}{2\sigma_e^2}\left[R \sum_{i=1}^{n} (x_i - \mu_i)^2 + \sum_{i=1}^{n} (y_i - \alpha - \beta\mu_i)^2\right]\right\}$$

$$\propto \frac{R^{n/2}}{\sigma_e^{2n}} \exp\left\{ -\frac{1}{2\sigma_e^2}\left[(R+\beta^2) \sum_{i=1}^{n} (\mu_i - \hat{\mu}_i)^2 + \frac{R}{R+\beta^2} \sum_{i=1}^{n} (y_i - \alpha - \beta x_i)^2 \right] \right\} \tag{23}$$

where $R = \sigma_e^2 / \sigma_d^2$, $\hat{\mu}_i = \frac{1}{R+\beta^2}[(Rx_i + \beta(y_i - \alpha))]$.

Note that $\hat{\mu}_i$ is exactly in the form of (8), the maximum likelihood estimator for μ_i.

As prior pdf for the parameters, we assume the following relationship:

$$p(\alpha, \beta, R, \sigma_e, \underset{\sim}{\mu}) \propto p_1(\alpha, \beta, R, \sigma_e)\, p_2(\underset{\sim}{\mu}) \tag{24}$$

where $p_2(\underset{\sim}{\mu})$ is given by (22) and $p_1(\alpha, \beta, R, \sigma_e)$, not yet specified, is discussed below. The posterior pdf is

$$p(\alpha, \beta, R, \sigma_e, \underset{\sim}{\mu} | \underset{\sim}{x}, \underset{\sim}{y}) \propto p_1(\alpha, \beta, R, \sigma_e)\, p_2(\underset{\sim}{\mu})\, L(\alpha, \beta, R, \sigma_e, \underset{\sim}{\mu} | \underset{\sim}{x}, \underset{\sim}{y}) \tag{25}$$

An approximate posterior pdf $p(\alpha, \beta, R, \sigma_e | \underset{\sim}{x}, \underset{\sim}{y})$ has been obtained by Lindley and El-Sayyad (1968). An asymptotic expression for the posterior distribution has been drived by Tiao and Zellner (1964). For obtaining the exact posterior pdf $p(\alpha, \beta, R, \sigma_e | \underset{\sim}{x}, \underset{\sim}{y})$, we take up the problem of integrating (25) with respect to the n elements of $\underset{\sim}{\mu}$. The factors of (25) involving $\underset{\sim}{\mu}$ are as follows:

$$\left[\sum_{i=1}^{n} (\mu_i - \bar{\mu})^2 \right]^{-(n-2)/2} \exp\left[-\frac{1}{2\sigma_e^2} (R+\beta^2) \sum_{i=1}^{n} (\mu_i - \hat{\mu}_i)^2 \right] \tag{26}$$

First, make the change of variables

$$z_i = \frac{(R+\beta^2)^{\frac{1}{2}}}{\sigma_e} (\mu_i - \hat{\mu}_i), \quad i = 1, 2, \ldots\ldots, n$$

or

$$\mu_i = \frac{\sigma_e}{(R+\beta^2)^{\frac{1}{2}}} (z_i + h_i), \quad i = 1. 2 \ldots, n$$

where

$$h_i = (R+\beta^2)^{\frac{1}{2}}\, \hat{\mu}_i / \sigma_e$$

Then

$$\prod_{i=1}^{n} d\mu_i = \sigma_e^{n} (R+\beta^2)^{-\frac{n}{2}} \prod_{i=1}^{n} dz_i$$

and (26) can be written in terms of $\underset{\sim}{z} = (z_1, \ldots, z_n)'$ as follows:

$$\frac{\sigma_e^2}{(R+\beta^2)} e^{-\frac{1}{2} \underset{\sim}{z}'\underset{\sim}{z}} [(\underset{\sim}{z}+\underset{\sim}{h})' M (\underset{\sim}{z}+\underset{\sim}{h})]^{-(n-2)/2} \tag{27}$$

where $\underset{\sim}{h} = (h_1, \ldots, h_n)'$, $M = I_n - \underset{\sim}{1}\underset{\sim}{1}'/n$ with I_n an $n \times n$ unit matrix and $\underset{\sim}{1}$ an $n \times 1$ column vector in which all elements are equal to one. On letting $\underset{\sim}{w} = \underset{\sim}{z} + \underset{\sim}{h}$, (27) becomes

$$\frac{\sigma_e^2}{(R+B^2)} e^{-\frac{1}{2}(\underset{\sim}{w}-\underset{\sim}{h})'(\underset{\sim}{w}-\underset{\sim}{h})} \left[\sum_{i=1}^{n} (w_i - \bar{w})^2\right]^{-(n-2)/2} \tag{28}$$

We can view the integration of (28) with respect to the elements of $\underset{\sim}{w}$ as the problem of finding the expectation of $[\sum (w_i - \bar{w})^2]^{-(n-2)/2}$ with the w_i's having a normal pdf. To get rid of $\bar{w}$ in the denominator, we use Helemert's transformation $\underset{\sim}{c} = B\underset{\sim}{w}$, which is

$$c_1 = (w_1 - w_2)/\sqrt{2}$$

$$c_2 = (w_1 + w_2 - 2w_3)/\sqrt{6}$$

$$c_3 = (w_1 + w_2 + w_3 - 3w_4)/\sqrt{12}$$

$$\vdots$$

$$c_{n-1} = [w_1 + \ldots + w_{n-1} - (n-1)w_n]/\sqrt{n(n-1)}$$

$$c_n = (w_1 + \ldots + w_n)/\sqrt{n}$$

Since the matrix B in the Helmert transformation is orthogonal, the Jacobian of the transformation is one, and from the properties of this transformation we have

$$\sum_{i=1}^{n} (w_i - \bar{w})^2 = \sum_{i=1}^{n-1} c_i^2$$

Further, since $E(\underset{\sim}{c}) = B\underset{\sim}{h}$, $\underset{\sim}{c} - E(\underset{\sim}{c}) = B(\underset{\sim}{w} - \underset{\sim}{h})$

Thus $\quad (\underset{\sim}{w}-\underset{\sim}{h})'(\underset{\sim}{w}-\underset{\sim}{h}) = [\underset{\sim}{c} - E(\underset{\sim}{c})]'(B^{-1})'B^{-1}[\underset{\sim}{c} - E(\underset{\sim}{c})]$

$$= [\underset{\sim}{c} - E(\underset{\sim}{c})]'[\underset{\sim}{c} - E(\underset{\sim}{c})]$$

Hence, we can express (28) as follows:

$$\frac{\sigma_e^2}{R+\beta^2}\left[\sum_{i=1}^{n} c_i^2\right]^{-(n-2)/2} \exp\left\{-\tfrac{1}{2}[\underset{\sim}{c} - E(\underset{\sim}{c})]'[\underset{\sim}{c} - E(\underset{\sim}{c})]\right\}$$

$$= \frac{\sigma_e^2}{(R+\beta^2)}\left[\sum_{i=1}^{n-1} c_i^2\right]^{-(n-2)/2} \exp\left[-\tfrac{1}{2}\sum_{i=1}^{n-1}(c_i - \theta_i)^2\right] \exp\left[-\tfrac{1}{2}(c_n - \theta_n)^2\right] \tag{29}$$

where $\theta_i = E(c_i)$, $i = 1, 2, \ldots, n$.

On integrating (29) with respecttto c_n, $-\infty$ to $+\infty$, we get a numerical constant. The integration with respect to the remaining c_i's is viewed as obtaining the expectation of $\Sigma_{i=1}^{n-1} c_i^2)^{-(n-2)/2}$ with the c_i's independent and normal, each with its its own mean θ_i. Under these conditions $u = \Sigma_{i=1}^{n-1} c_i^2$ has the following noncentral chisquare pdf:

$$(2^p\pi)^{-\frac{1}{2}} e^{-\frac{1}{2}(u+\lambda)} u^{\frac{1}{2}p-1} \sum_{k=0}^{\infty} \frac{\lambda^k u^k \Gamma(k+\frac{1}{2})}{(2k)!\,\Gamma(k+\frac{1}{2}p)} \tag{30}$$

where $p = n-1$, and the noncentrality parameter λ is given by

$$\lambda = \sum_{i=1}^{n-1} \theta_i^2 = \sum_{i=1}^{n} (h_i - \overline{h})^2$$

Then on multiplying (30)by $u^{-(n-2)/2}$, the integral of interest is proportional to

$$\frac{\sigma_e^2}{R+\beta^2} e^{-\frac{1}{2}\lambda} \int_0^{\infty} \sum_{k=0}^{\infty} \frac{\lambda^k u^{k-\frac{1}{2}}}{(2k)!} e^{-\frac{1}{2}u} \frac{\Gamma(k+\frac{1}{2})}{\Gamma(k+\frac{1}{2}p)} du \tag{31}$$

Since $\int_0^{\infty} u^{k-\frac{1}{2}} e^{-\frac{1}{2}u} du = 2^{k+\frac{1}{2}} \Gamma(k+\frac{1}{2})$,

the integral in (31) is proportional to

$$\frac{\sigma_e^2}{R+\beta^2} e^{-\frac{1}{2}\lambda} \sum_{k=0}^{\infty} \frac{\lambda^k 2^k}{(2k)!} \frac{[\Gamma(k+\frac{1}{2})]^2}{\Gamma(k+\frac{1}{2}p)} \tag{32}$$

from the following properties of the gamma function

$$\Gamma(k+\tfrac{1}{2}) = (2k)!\sqrt{\pi}\,\Gamma(k+1)/2^{2k}$$

$$\Gamma(k+1) = k!$$

We can rewrite (32) as follows:

$$\frac{\sigma_e^2}{R+\beta^2} e^{-\frac{1}{2}\lambda} \sum_{k=0}^{\infty} \left(\frac{\lambda}{2}\right)^k \frac{\Gamma(k+\frac{1}{2})}{k!\,\Gamma(k+\frac{1}{2}p)} \tag{33}$$

where $\lambda = \sum_{i=1}^{n} (h_i - \overline{h})^2 = \frac{R+\beta^2}{\sigma_e^2} \sum_{i=1}^{n} (\hat{\mu}_i - \overline{\hat{\mu}})^2$

$$\bar{\hat{\mu}} = \frac{1}{n} \sum_{i=1}^{n} \hat{\mu}_i$$

Therefore, on integrating (25) with respect to the elements of $\underset{\sim}{\mu}$, we obtain

$$p(\alpha, \beta, R, \sigma_e \mid \underset{\sim}{x}, \underset{\sim}{y}) \propto p_1(\alpha, \beta, R, \sigma_e) \frac{R^{n/2}}{\sigma_e^{2n-2}(R+\beta^2)}$$

$$\times \exp\left[-\frac{1}{2\sigma_e^2} \frac{R}{R+\beta^2} \sum_{i=1}^{n} (y_i - \alpha - \beta x_i)^2\right]$$

$$\times e^{-\frac{1}{2}\lambda} \sum_{k=0}^{\infty} \left(\frac{\lambda}{2}\right)^k \frac{\Gamma(K+\frac{1}{2})}{k!\,I(k+\frac{1}{2}p)} \tag{34}$$

where $\lambda = \dfrac{R+\beta^2}{\sigma_e^2} \displaystyle\sum_{i=1}^{n} (\hat{\mu}_i - \bar{\hat{\mu}})^2$

$$= \frac{n}{\sigma_e^2 (R+\beta^2)} (R^2 m_{xx} + \beta^2 m_{yy} + 2R\beta m_{xy})$$

$$m_{xx} = \frac{1}{n} \sum_{i=1}^{n} (x_i - \bar{x})^2, \qquad m_{yy} = \frac{1}{n} \sum_{i=1}^{n} (y_i - \bar{y})^2$$

$$m_{xy} = \frac{1}{n} \sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y})$$

As regards the prior pdf $p_1(\alpha, \beta, R, \sigma_e)$ in (34), we shall analyze two cases below:

(1) α, β, R and σ_e are assumed to be independently distributed.

If our information about α and σ_e is vague, then the prior pdf can be taken as

$$p_1(\alpha, \beta, R, \sigma_e) \propto \frac{g_1(\beta)\, g_2(R)}{\sigma_e} \qquad -\infty < \alpha < +\infty \quad 0 < \sigma_e < +\infty \tag{35}$$

with $g_1(\beta)$ and $g_2(R)$ still unspecified prior pdf's for β and R, respectively.

On substituting from (35) into (34) and integrating with respect to α, the resulting posterior pdf is

$$p(\beta, R, \sigma_e \mid x, y) \propto \frac{g_1(\beta)\, g_2(R)\, R^{(n-2)/2}}{\sigma_e^{2n-2} (R+\beta^2)^{\frac{1}{2}}}$$

$$\times \exp\left\{-\frac{1}{2\sigma_e^2}\ \frac{R}{R+\beta^2}\sum_{i=1}^{n}[\,y_i-\bar{y}-\beta(x_i-\bar{x})\,]^2\right\}$$

$$\times\, e^{-\lambda_1/2\sigma_e^2}\sum_{k=0}^{\infty}\left(\frac{\lambda_1}{2\sigma_e^2}\right)^k \frac{\Gamma(k+\frac{1}{2})}{k!\,\Gamma(k+\frac{1}{2}p)} \qquad (36)$$

where $\lambda_1 = \lambda\,\sigma_e^{\,2}$.

Furthermore, on integrating (36) termwise with respect to σ_e, the result is

$$p(\beta,\ R\,|\,\underset{\sim}{x},\ \underset{\sim}{y}) \propto \frac{g_1(\beta)\, g_2(R)\ \ R^{(n-1)/2}}{(R+\beta^2)^{\frac{1}{2}}}\ \frac{1}{A^{n-3/2}}\sum_{k=0}^{\infty}\left(\frac{\lambda_1}{A}\right)^k B_k \qquad (37)$$

where $A = \dfrac{R}{R+\beta^2}\displaystyle\sum_{i=1}^{n}[(y_i-\bar{y}-\beta(x_i-\bar{x})^2] + \lambda_1$

$$B_k = \frac{\Gamma(k+n-3/2)\ \ \Gamma(k+\frac{1}{2})}{k!\ \Gamma[k+(n-1)/2]}$$

By straightforward algebra it is the case that

$$A = n\,(m_{yy} + Rm_{xx})$$

$$\frac{\lambda_1}{A} = \frac{R^2 m_{xx} + \beta^2 m_{yy} + 2R\beta m_{xy}}{(m_{yy}+Rm_{xx})\,(R+\beta^2)}$$

Thus (37) can be expressed as

$$p(\beta,\ R\,|\,\underset{\sim}{x},\ \underset{\sim}{y})\ \frac{g_1(\beta)\ g_2(R)}{(R+\beta^2)^{\frac{1}{2}}}\ \frac{R^{(n-1)/2}}{(m_{yy}+Rm_{xx})^{n-3/2}}$$

$$\times\sum_{k=0}^{\infty}\left[\frac{R^2 m_{xx} + \beta^2 m_{yy} + 2R\beta m_{xy}}{(m_{yy}+Rm_{xx})\,(R+\beta^2)}\right]^k B_k \qquad (38)$$

To use (38) in practice prior pdf's for β and R must be assigned. With respect to the prior pdf for β, we can assign a beta pdf of the following form:

$$p(z\,|\,a,\ b) = \frac{1}{B(a,\ b)}\, z^{a-1}(1-z)^{b-1},\quad \begin{matrix}0<a,\ b\\ 0\leqq z\leqq 1\end{matrix} \qquad (39)$$

where $z = (\beta - \beta_L) / (\beta_U - \beta_L)$, β_L, β_U, a and b are prior parameters to be assigned by an investigator, and B(a, b) denotes the beta function. If a = b = 1, (39) gives us a uniform pdf for β. With respect to R, the prior pdf might be taken in the following inverted gamma form:

$$g_2(R \mid v_o, s_o) \propto R^{-(v_o+1)} \exp\left(-\frac{v_o s_o^2}{2R^2}\right), \quad \begin{array}{l} 0<R<\infty \\ 0<v_o, s_o \end{array} \tag{40}$$

where v_o and s_o are prior parameters.

Substituting from (39) and (40) into (38), we have a bivariate posterior pdf which can be analysed using bivariate numerical integration techniques. Note that the term λ_1/A in (38), for given R, has a maximum at $\beta = \hat{\beta}$, where $\hat{\beta}$ is the maximum likelihood estimate, that is, for given R,

$$\frac{d}{d\beta}\left(\frac{\lambda_1}{A}\right) = \frac{1}{m_{yy} + Rm_{xx}}\left[\frac{2(\beta m_{yy} + Rm_{xy})}{R + \beta^2} - \frac{2\beta(R^2 m_{xx} + \beta^2 m_{yy} + 2R\beta m_{xy})}{(R + \beta^2)^2}\right]$$

On setting this derivative equal to zero, the necessary condition for a maximum is

$$\beta^2 m_{xy} + \beta\,(Rm_{xx} - m_{yy}) - Rm_{xy} = 0$$

A solution of the quadratic equation is identical to (6). Thus for given R, (38) will have a conditional modal value for β close to the maximum likelihood estimate.

(2) α, β, σ_d and σ_e are assumed to be independently distributed.

If we assume that our prior information about independent parameters σ_d and σ_e can be represented by inverted gamma pdf's:

$$\begin{aligned} p(\sigma_e \mid v_d, s_d) &\propto \sigma_d^{-(v_d+1)} \exp(-v_d s_d^2 / 2\sigma_d^2), \quad 0<\sigma_d<\infty \\ p(\sigma_e \mid v_e, s_e) &\propto \sigma_e^{-(v_e+1)} \exp(-v_e s_e^2 / 2\sigma_e^2), \quad 0<\sigma_e<\infty \end{aligned} \tag{41}$$

where v_d, s_d, v_e, s_e are prior parameters. Now, on transforming to $R = \sigma_e^2/\sigma_d^2$, we have the joint prior pdf of R and σ_e as follows:

$$g_3(R, \sigma_e) \propto \frac{R^{(v_d-2)/2}}{\sigma_e^{v_d+v_e+1}} \exp\left[-\frac{1}{2\sigma_e^2}(Rv_d s_d^2 + v_e s_e^2)\right] \tag{42}$$

It is seen that the conditional prior pdf for σ_e, given R, is in the inverted gamma form, and the marginal pdf for Rs_d^2/s_e^2 is in the form of a Fisher-Snedecor F pdf with v_d and v_e degrees of freedom. Then, if in place of (35) we use the prior pdf

$$P_1(\alpha, \beta, R, \sigma_e) \propto g_1(\beta)\, g_3(R, \sigma_e) \tag{43}$$

where $g_3(R, \sigma_e)$ as shown in (42) and $g_1(\beta)$ not yet specified, we can substitute from (43) into (36) and perform integrations with respect to α and σ_e in much the same way as shown above. The resulting posterior pdf for β and R is given by

$$p(\beta, R \mid \underset{\sim}{x}, \underset{\sim}{y}) \propto \frac{g_1(\beta)\, R^{(n+v_d-3)/2}}{(R+\beta^2)^{\frac{1}{2}}}\, G^{-n-(v-3)/2} \sum_{k=0}^{\infty} \left(\frac{\lambda_1}{G}\right)^k B_k' \tag{44}$$

where $v = v_d + v_e$,

$$B_k' = \frac{\Gamma[k+n+(v-3)/2]\, \Gamma(k+\frac{1}{2})}{k!\, \Gamma[k+(n-1)/2]}$$

$$G = A + Rv_d s_d^2 + v_e s_e^2 = n\left[\frac{nm_{yy} + v_e s_e^2}{n} + \frac{R(nm_{xx} + v_d s_d^2)}{n}\right]$$

$$\frac{\lambda_1}{G} = \frac{R^2 m_{xx} + \beta^2 m_{yy} + 2R\beta m_{xy}}{(R+\beta^2)[(nm_{yy} + v_e s_e^2)/n + R(nm_{xx} + v_d s_d^2)/n]}$$

where λ_1 and A have been defined in connection with (36) and (37). Thus, we can write the joint posterior pdf for β and R as follows:

$$p(\beta, R \mid \underset{\sim}{x}, \underset{\sim}{y}) \propto \frac{g_1(\beta)}{(R+\beta^2)^{\frac{1}{2}}}\, \frac{R^{(n+v_d-3)/2}}{(a_o + Ra_1)^{(2n+v-3)/2}}$$

$$\times \sum_{k=0}^{\infty} \left[\frac{R^2 m_{xx} + \beta^2 m_{yy} + 2R\beta m_{xy}}{(R+\beta^2)(a_o + Ra_1)}\right]^k B_k \tag{45}$$

where $a_o = (nm_{yy} + v_e s_e^2)/n$, $a_1 = (nm_{xx} + v_d s_d^2)/n$.

Note that in(45) the factors

$$R^{(n+v_d-3)/2} / (a_o + Ra_1)^{(2n+v-3)/2 + k}, \quad k = 0, 1, 2, \ldots.$$

are in the form of F pdf's with v_d+n-1 and $v_e+n-2+2k$ degrees of freedom. As mentioned above, for given R, the quantity $(R^2m_{xx}+\beta^2m_{yy}+2R\beta m_{xy})/(R+\beta^2)$ has a maximum at the maximum likelihood estimate of β. Thus for given R, (45) will have a conditional modal value for β close to the maximum likelihood estimate.

Ⅳ. Conclusion and Summary

The linear model for n pairs of observations (x_i, y_i), i=1, 2, ... , n, considered in this study is defined as

$$x_i = \mu_i + d_i \qquad i = 1, 2, \ldots, n$$
$$y_i = \alpha + \beta\mu_i + e_i$$

where the errors d_i's and e_i's are independently and normally distributed with

$$E(d_i) = 0, \qquad E(e_i) = 0$$
$$E(d_i^2) = \sigma_d^2, \qquad E(e_i^2) = \sigma_e^2$$
$$E(d_ie_i) = 0.$$

The maximum likelihood approach to the estimation of the parameters in the model is compared with the Bayesian approach.

(1) If all the parameters in the model are assumed to be unknown, then a maximum of the likelihood function does not exist in the admissible region of the parameter space. If the ratio of the two error variances, $R=\sigma_e^2/\sigma_d^2$, is known exactly, then the maximum likelihood estimators for α and β are consistent whereas the estimator for σ_d^2 is not. Further, if the μ_i's are assumed to be independent of the d_i's and e_i's and identically, normally and independently distributed with mean τ and variance σ^2, then under the assumption that $\alpha = 0$, the meaningless variance estimates for σ_d^2, σ_e^2 and σ^2 can be obtained and the estimate of β falls in the inadmissible region of the parameter space. If, rather α, $R = \sigma_e^2/\sigma_d^2$ is known, then the maximum likelihood estimate of β is just in precisely the same form as in the case where the μ_i's are unknown parameters. If the μ_i's are assumed to be uniformly and independently distributed, then the maximum likelihood estimates for α and β are just the simple least squares estimates obtained from regressing $\underset{\sim}{y}$ on $\underset{\sim}{x}$. It is seen that the maximum likelihood estimates for the parameters in the model are vitally dependent on the assumption about the pdf for the μ_i's.

(2) In the Bayesian approach, there is no need to assume the ratio $R=\sigma_e^2/\sigma_d^2$ to be know exactly in order to make inferences about parameters of interest. If we assume a priori that the μ_i's, α, β, $\ln\sigma_d$, $\ln\sigma_e$ are uniformly and independently distributed with the prior pdf $p(\underset{\sim}{\mu}, \alpha, \beta, \sigma_d, \sigma_e) \propto (\sigma_d\sigma_e)^{-1}$, then α and β have a posterior pdf in the bivariate Student t form with mean α and $\hat{\beta}$, precisely the least squares quantity obtained from regressing $\underset{\sim}{y}$ on $\underset{\sim}{x}$. If the μ_i's are assumed, a priori, to be normally and independently distributed with mean τ and variance σ^2 and assume further that τ and σ are unknown and distributed with a priori pdf $p(\tau, \sigma) \propto$ const. then the exact joint posterior pdf's for β and R, $p(\beta, R \mid \underset{\sim}{x}, \underset{\sim}{y})$ are obtained under two different assumptions about the prior pdf for α, β, R, and σ_e:

(i) $p_1(\alpha, \beta, R, \sigma_e) \propto g_1(\beta)\, g_2(R)/\sigma_e$

where $g_1(\beta)$ and $g_2(R)$ are unspecified prior pdf's for β and R, respectively.

(ii) $p_1(\alpha, \beta, R, \sigma_e) \propto g_1(\beta)\, g_3(R, \sigma_e)$

where $g_3(R, \sigma_e)$ is the joint prior pdf of R and σ_e obtained from the independent parameters σ_d and σ_e having, a priori, the inverted gamma distribution.

It is found that both two joint posterior pdf's for β and R have, for given R, the same conditional model value for β close to the maximum likelihood estimate $\hat{\beta}$ which is obtained by the maximum likelihood approach under the assumption that R is known.

References

1. James, W. and Stein, C. M. (1961): "Estimation with quadratic loss" in Neyman, J., Ed. Proceeding of Fourth Berkeley Symposium on Mathematical Statistics and Probability, Vol. 1, Berkeley: University of California Press, 361-379.
2. Kendall, M.G. and Stuart, A. (1961): The Advanced Theory of Statistics, Vol.2, New York: Hafner, p. 61.
3. Lindley, D.V. and El-Sayyed, G.M. (1968): The Bayesian estimation of a linear functional relationship. Journal of Royal Statistical Society, Series B, 30, 190-202.
4. Madansky, A. (1959): The fitting of straight lines when both variables are subject to error. Journal of American Statistical Association 54, 173-205.
5. Neyman, J. and Scott, E.L. (1951): On certain methods of estimating the linear structural relation. Annals of Mathematical Statistics, 22, 352-361.

6. Stein, C. M. (1956): "Inadmissibility of the usual estimatior for the mean of a multivariate normal distribution" in Neyman, J. Ed. Proc. Third Berkeley Symp. Math. Statist. Probab., Vol. 1, Berkeley: University of California Press, 197-206.

7. Stein, C, M. (1962): Confidence sets for the mean of a multivariate normal distribution. Journal of Royal Statistical Society, Series B, 24, 265-285.

8. Tiao, G. C. and Zeliner, A. (1964): Bayes's theorem and the use of prior knowledge in regression analysis. Biometrika, 51, 219-230.