© Journal of the American Statistical Association
March 1970, Volume 65, Number 329
Theory and Methods Section

A Bayesian Look at Inverse Linear Regression

BRUCE HOADLEY*

The model considered in this paper is simple linear regression ($Ey_i = \beta_1 + \beta_2 x_i$, $i = 1, \cdots, n$), and the problem is to make statistical inferences about an unknown value of x corresponding to one or more additional observed values of y. The maximum likelihood estimator $\hat{x}$ of x and the classical $(1-\alpha)$ 100% confidence set S for x have some undesirable properties. For example, $\hat{x}$ has infinite mean square error and $P\{S = (-\infty, +\infty)\} > 0$. The purpose of this paper is to demonstrate that insight and understanding, as well as a useful class of solutions, can be obtained by looking at the problem from a Bayesian point of view.

A result which follows from a general Bayes solution is that the inverse estimator [4] is Bayes with respect to a particular informative prior.

1. INTRODUCTION AND DISCUSSION

The *inverse linear regression problem* can be stated formally as follows: observations take the form

$$\begin{aligned} y_{1i} &= \beta_1 + \beta_2 x_i + \epsilon_{1i} \qquad i = 1, \cdots, n \\ y_{2j} &= \beta_1 + \beta_2 x + \epsilon_{2j} \qquad j = 1, \cdots, m, \end{aligned} \tag{1.1}$$

where the ϵ_{1i}'s and ϵ_{2j}'s are mutually independent and identically distributed as $N(0, \sigma^2)$. It is assumed that $x_1, \cdots, x_n$ are known constants, and that β_1, β_2, σ^2, and x are unknown. The problem is to make statistical inferences about x based on $y_{11}, \cdots, y_{1n}, y_{21}, \cdots, y_{2m}$. Without loss of generality, the x_i's are chosen so that

$$\sum_i x_i = 0, \qquad \left[\sum_i x_i^2\right]/n = 1. \tag{1.2}$$

This problem is sometimes called the *linear calibration problem*, where, for example, the x_i's might be known weights, and the y_{1i}'s the corresponding readings off the scale being calibrated. If someone wanted to weigh an object with unknown weight x, he might take m different readings $y_{21}, \cdots, y_{2m}$ from the scale, and then use the y_{1i}'s, y_{2j}'s, and the x_i's to estimate x. The problem is also called the *discrimination problem* [5, p. 338], or *the reverse of the prediction problem*, and has applications to problems of biological assay. For example, $x_1, \cdots, x_n$ might be the known concentrations of a vitamin in n batches of chicken feed and $y_{11}, \cdots, y_{1n}$ might be the corresponding observed weight gains of n young chicks. One way to estimate the vitamin concentration, x, in a new batch of feed would be to observe weight gains $y_{21}, \cdots, y_{2m}$ of m additional chicks, and then use the model in (1.1).

* Bruce Hoadley is a member of the Statistics Department, Bell Telephone Laboratories, Holmdel, N. J. His previous publications have appeared in the *Annals of Mathematical Statistics*, *Biometrika*, and the *Journal of the American Statistical Association*. The author wishes to thank Brian Joiner and David Hogben, National Bureau of Standards, for stimulating discussions. He is also indebted to Dan Relles and Andrew Kalotay, Bell Telephone Laboratories, who helped with some of the technical, as well as philosophical, points in the article.

The discussion in the statistical literature on how to make inferences about x can be characterized by disagreement and confusion. The purpose of this paper is to view some of this discussion in a Bayesian light and then show that a Bayesian approach to the problem provides valuable insight as well as a usable class of solutions.

Before proceeding, some notation must be introduced and some elementary facts stated. Vectors and matrices shall be denoted by lower and upper case boldface letters, respectively. So, $\boldsymbol{y}$ is a vector and $\boldsymbol{X}$ is a matrix. The maximum likelihood and/or least squares estimators of β_1 and β_2 based on $\boldsymbol{y}_1$ are

$$\begin{aligned} \hat{\beta}_1 &= \bar{y}_1 \\ \hat{\beta}_2 &= \left[\sum_i y_{1i} x_i \right] / n. \end{aligned} \tag{1.3}$$

The classical unbiased estimators of σ^2 based on $\boldsymbol{y}_1$ alone, $\boldsymbol{y}_2$ alone, and both $\boldsymbol{y}_1$ and $\boldsymbol{y}_2$ are

$$\begin{aligned} v_1 &= \left[\sum_i (y_{1i} - \hat{\beta}_1 - \hat{\beta}_2 x_i)^2 \right] / \nu_1 \\ v_2 &= \left[\sum_j (y_{2j} - \bar{y}_2)^2 \right] / \nu_2 \\ v &= [\nu_1 v_1 + \nu_2 v_2]/\nu, \end{aligned} \tag{1.4}$$

respectively, where

$$\begin{aligned} \nu_1 &= n - 2 \\ \nu_2 &= m - 1 \\ \nu &= \nu_1' + \nu_2. \end{aligned} \tag{1.5}$$

The solution to the estimation problem most widely recommended in the literature[1] is the maximum likelihood estimator (MLE). It can be shown [5, p. 339] that the MLE is

$$\begin{aligned} \hat{x} &= [\bar{y}_2 - \hat{\beta}_1]/\hat{\beta}_2 \\ &= [\bar{y}_2 - \bar{y}_1]/\hat{\beta}_2. \end{aligned} \tag{1.6}$$

One cannot judge this "classical" solution by the familiar "classical" criterion of mean-square error (MSE) because

$$E[(\hat{x} - x)^2 \mid \beta_1, \beta_2, \sigma^2, x] = +\infty. \tag{1.7}$$

In a Monte Carlo experiment, with $m = 1$, Krutchkoff [4] showed that this has practical significance; because, for $|\beta_2/\sigma| < 10$, the empirical efficiency (measured by the ratio of average squared errors) of the *inverse estimator* (defined below) relative to a truncated version of $\hat{x}$ (truncation was used to prevent $\hat{\beta}_2$ from getting too close to zero) was noticeably larger than 1 for a wide variety of conditions. For example, when $\beta_2/\sigma = 5$, the empirical efficiency was near

[1] For references, see [4].

1.25 in most cases considered. The inverse estimator is

$$\begin{aligned} \hat{x}_I &= \hat{\gamma} + \hat{\delta}\bar{y}_2 \\ &= [\bar{y}_2 - \bar{y}_1]\hat{\delta}, \end{aligned} \tag{1.8}$$

where

$$\begin{aligned} \hat{\delta} &= \left[\sum_i y_{1i}x_i\right] \Big/ \left[\sum_i (y_{1i} - \bar{y}_1)^2\right] \\ \hat{\gamma} &= -\hat{\delta}\bar{y}_1 \end{aligned} \tag{1.9}$$

are the least squares estimators of the slope and intercept when the x_i's are formally regressed on the y_{1i}'s.

Although it is true that the MSE of $\hat{x}_I$ is finite, Williams [8] remarked that this does not prove very much. He went on to show that if σ^2 and the sign of β_2 are known, then the unique unbiased estimator of x has infinite variance. In fact, he stated that, " $\cdots$ since the classical, the unbiased, or indeed any estimator that could be derived in a theoretically justifiable manner all have infinite variances, the fact that Krutchkoff's estimator has finite variance seems to be of little account."

Notwithstanding this argument, I feel that $\hat{x}$ is unsatisfactory from a point of view which is independent of MSE considerations. Let

$$F = n\hat{\beta}_2^2/v; \tag{1.10}$$

this is the F-statistic often used for testing the hypothesis that $\beta_2 = 0$. Intuitively, if F is much larger than $F_{\alpha;1,\nu}$ (the upper α point of the F-distribution with 1 and ν degrees of freedom, which is the distribution of F when $\beta_2 = 0$) then $\hat{x}$ is fairly precise, but if the opposite is true, then $\hat{x}$ is very imprecise (this can be seen clearly by plotting some data). In other words, the data contain information about the precision of $\hat{x}$; so, it seems reasonable to have some way of giving $\hat{x}$ less weight when it is known to be unreliable. This is precisely what a Bayes estimator does.

The most interesting by-product of the Bayes solutions proposed in this paper is a characterization of the inverse estimator $\hat{x}_I$. For $m = 1$, it is shown in Section 3 that if x has a t prior density with $n-3$ (the same n as in (1.1)) degrees of freedom, mean 0, and scale parameter $[(n+1)/(n-3)]^{1/2}$, then, *a posteriori*, x has a t density with $n-2$ degrees of freedom and mean $\hat{x}_I$. So the inverse estimator is Bayes with respect to squared error loss and a particular informative prior distribution for x. Therefore, from a Bayesian point of view, $\hat{x}_I$ is justifiable when the above peculiar prior distribution approximately quantifies the available prior knowledge of x. It appears that Williams [8, p. 191, l. 33] was wrong when he stated implicitly that the inverse estimator cannot be derived in a theoretically justifiable manner. The above characterization provides $\hat{x}_I$ with some support since in practice it may be highly probable that x is within the range of $x_1, \cdots, x_n$. However, I would recommend a more careful selection of a prior distribution on x.

The classical $(1-\alpha)$ 100% confidence set S for x [5, p. 339] also possesses inherent difficulties. If $m = 1$, the confidence set is derived from the fact that

$$\frac{n^{1/2}\hat{\beta}_2(\hat{x}-x)}{[v(n+1+x^2)]^{1/2}} \tag{1.11}$$

has a t distribution with $n-2$ degrees of freedom. From (1.11) it follows that

$$S=\begin{cases}\{x\colon x_L\le x\le x_U\} & \text{if } F>F_{\alpha;1,n-2}\\ \{x\colon x\le x_L\}\cup\{x\colon x\ge x_U\} & \text{if } \left[\dfrac{n+1}{n+1+\hat{x}^2}\right]F_{\alpha;1,n-2}\le F<F_{\alpha;1,n-2}\\ (-\infty,+\infty) & \text{if } F<\left[\dfrac{n+1}{n+1+\hat{x}^2}\right]F_{\alpha;1,n-2},\end{cases} \tag{1.12}$$

where F is defined in (1.10) and x_L and x_U are equal to

$$\frac{F\hat{x}}{(F-F_{\alpha;1,n-2})}\pm\frac{\{F_{\alpha;1,n-2}[(n+1)(F-F_{\alpha;1,n-2})+F\hat{x}^2]\}^{1/2}}{(F-F_{\alpha;1,n-2})}, \tag{1.13}$$

with $x_L<x_U$ (see Figure 1 for a graphical display of S). So Williams' [8, p. 192] suggestion to use a confidence interval estimator instead of a point estimator would not be very helpful if

$$F<[(n+1)/(n+1+\hat{x}^2)]F_{\alpha;1,n-2}.$$

The set S was also obtained by Fieller [3] using a fiducial argument, and he said that one should intuitively expect S to have infinite length whenever $F<F_{\alpha;1,n-2}$, because then $\hat{\beta}_2$ is not significantly different from zero. This statement may tempt one to conclude that in such a case the data provide no information about x. A Bayesian would agree that the data provide no information about x if his prior and posterior distributions of x are the same. For the Bayesian model of Section 2, they are different whenever $F>0$; i.e., with probability 1, the data will provide information about x. In Figure 4, the posterior for Simulation 1 is an example of a posterior in the case of a diffuse prior and $F<F_{\alpha;1,n-2}$. Of course, the closer F is to 0, the more diffuse the posterior would become.

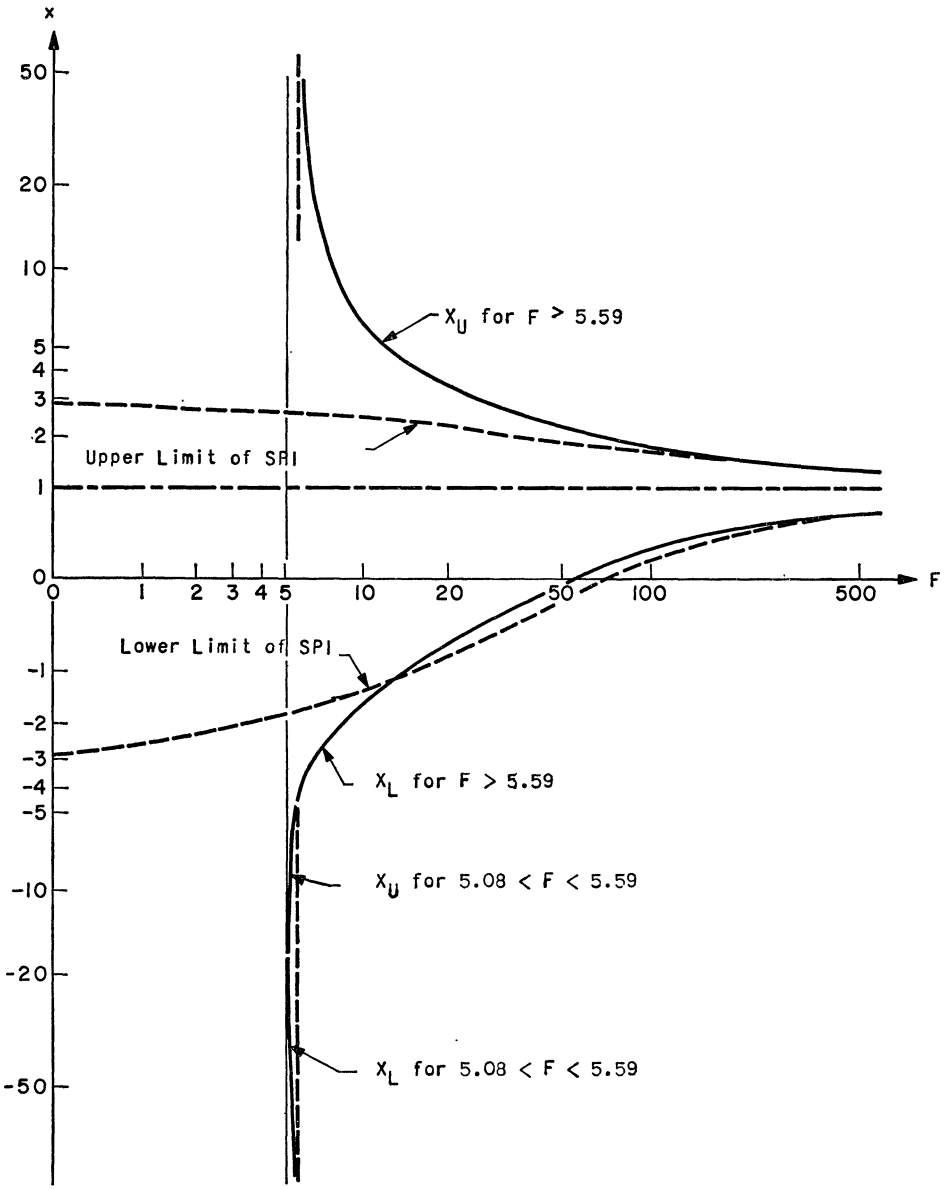
In any case, $1-\alpha$ does not measure anyone's "confidence" that S contains x, particularly when $S=(-\infty,+\infty)$. Of course this paradox is explained by the fact that $1-\alpha$ is associated with a property of the random set S, not with any particular realization of S. The user is left with the following relevant questions unanswered:

1. What measure of confidence can be associated with the particular realization of S at hand?
2. Using the data at hand, how does one construct an interval of finite length which contains x with $(1-\alpha)$ 100% "confidence"?

A posterior distribution on x would provide an answer to both questions. In Figure 1 the classical confidence set is compared with the *shortest posterior interval* (SPI) obtained from the aforementioned posterior distribution whose mean is $\hat{x}_I$.

Fieller [3, p. 176] also showed that the inverse linear regression problem can be reduced to considering the ratio of two means. Let $\eta=\beta_2 x$, $\xi=\beta_2$, $\hat{\eta}=\bar{y}_2-\hat{\beta}_1$,

Figure 1. COMPARISON OF 95% CONFIDENCE SET AND 95% SPI FOR $n=9$, $\hat{x}=1$



and $\hat{\xi}=\hat{\beta}_2$. Then $x=\eta/\xi$, and $(\hat{\eta}, \hat{\xi})$ has a multivariate normal distribution with mean and covariance equal to

$$\begin{matrix} & (\eta, \xi) \\ \sigma^2 & \begin{bmatrix} (1/m+1/n) & 0 \\ 0 & 1/n \end{bmatrix} \end{matrix} \tag{1.14}$$

respectively; and $(\hat{\eta}, \hat{\xi})$ is statistically independent of v. Now one can use either Fieller's [3] or Creasy's [1] method to obtain an interval estimator for x.

However, from a Bayesian point of view, there are serious flaws in the preceding ideas. First of all, the minimal sufficient statistic for $(\beta_1, \beta_2, \sigma^2, x)$ is $(\hat{\beta}_1, \hat{\beta}_2, v, \bar{y}_2)$; so one cannot be sure, *a priori*, that the posterior distribution of x will depend only on $(\hat{\eta}, \hat{\xi}, v)$ (although this turns out to be the case in my formulation). In other words, one is not guaranteed, *a priori*, that the reduction to $(\hat{\eta}, \hat{\xi}, v)$ entails no loss of information about x. Secondly, Creasy computed a fiducial distribution for x by first computing conditional fiducial distributions for η and ξ given σ^2; then treating η and ξ as independent to obtain a fiducial distribution of x given σ^2; and then integrating out σ^2 wrt its fiducial distribution. Savage [7, p. 24] pointed out that his theory of precise measurement (i.e., the Bayesian approach) provides approximately the same solution if η and ξ are independent normal means. However, if one assumes that, *a priori*, x and β_2 are independent (which is reasonable, since β_2 is a property of the instrument and x is a property of the object being measured), then, *a posteriori*, $\eta = \beta_2 x$ and $\xi = \beta_2$ will not be independent, and Creasy's solution is not applicable.

Before developing a Bayes solution to the inverse linear regression problem, it should be mentioned that Dunsmore [2] derived a Bayes solution to a similar problem where $m = 1$, and (x, y_2), (x_i, y_{1i}) $i = 1, \cdots, n$ are assumed to be independent observations from a bivariate normal distribution. He obtained $\hat{x}_I$ as the conditional mean of x given y_2, (x_i, y_{1i}) $i = 1, \cdots, n$. This is not too surprising since, for the bivariate model, the estimation of x is really a prediction problem (as opposed to a reverse prediction problem) and classical principles can be used to derive $\hat{x}_I$ as a predictor for x.

2. A BAYES SOLUTION

First, more notation is needed. If θ is an unknown parameter, then $p(\theta)$ and $p(\theta|\text{ data})$ denote the prior and posterior density of θ, respectively. The likelihood function is denoted by $l(\theta|\text{ data})$; and the conditional density of y given w, where y is an observable variable, is denoted by $f(y|\, w)$. All parameters and data variables are sometimes considered as constants and sometimes as random variables. It is not necessary to complicate the notation by distinguishing between the two.

Examples of symbolic distributional statements are:

$$h \sim \frac{1}{v}\,[\chi_\nu^2/\nu] \qquad \{\beta \mid y\} \sim t_\nu(m, v). \tag{2.1}$$

The first means that vvh is distributed as χ^2 with ν degrees of freedom; and the second means that conditional on y, β has the same distribution as $m + t_\nu\sqrt{v}$, where t_ν has a t distribution with ν degrees of freedom.

A useful lemma [6, p. 235] is

Lemma 1. If $\{y|\sigma^2\} \sim N(m, \sigma^2/n)$ *and* $1/\sigma^2 \sim (1/v)[\chi_\nu^2/\nu]$, *then* $y \sim t_\nu(m, v/n)$.

The first step in a Bayesian approach is to select some prior distribution.

Here it is assumed that $(\beta, \ln \sigma)$ has a uniform distribution; i.e.,

$$p(\beta, \sigma^2) \propto 1/\sigma^2. \tag{2.2}$$

This is the improper "noninformative" prior which is often selected in regression problems, and it can be viewed as an approximate representation of vagueness. With no additional technical difficulty the analysis could be carried out under the assumption of the natural conjugate family of priors for (β, σ^2) (Raiffa and Schlaifer, [6]); however, assumption (2.2) simplifies the formulas considerably, and does not detract from the main theme of the paper. The prior density of x is arbitrary and will be denoted by $p(x)$.

The posterior density of x is given by the following theorem:

Theorem 1. Suppose that, a priori, x is independent of (β, σ^2), and the prior distribution of (β, σ^2) is specified by (2.2). *Then the posterior density of x is given by*

$$p(x \mid y_1, y_2) \propto p(x)L(x), \tag{2.3}$$

where

$$L(x) = \frac{[1 + n/m + x^2]^{\nu/2}}{[1 + n/m + R\hat{x}^2 + (F/\nu + 1)(x - R\hat{x})^2]^{(\nu+1)/2}} \tag{2.4}$$
$$R = F/(F + \nu).$$

Proof. There are at least two ways to derive this result. The one suggested by a referee is to use Bayes theorem to get the joint posterior distribution of (x, σ^2, β), and then integrate out β and σ^2 in that order. I find this approach algebraically tedious and prefer a more deductive approach which seems to provide more insight.

The following list of facts is an outline of the derivation:

(A) $$p(x \mid y_1, y_2) \propto p(x)f(\bar{y}_2 \mid x, y_1, v_2).$$

(B) $$\left\{\begin{bmatrix}\beta_1\\ \beta_2\\ \bar{\epsilon}_2\end{bmatrix} \middle| \sigma^2, x, y_1, v_2\right\} \sim N\left(\begin{bmatrix}\hat{\beta}_1\\ \hat{\beta}_2\\ 0\end{bmatrix}, \sigma^2\begin{bmatrix}1/n & 0 & 0\\ 0 & 1/n & 0\\ 0 & 0 & 1/m\end{bmatrix}\right)$$

(C) $$\{\bar{y}_2 \mid \sigma^2, x, y_1, v_2\} \sim N(\hat{\beta}_1 + \hat{\beta}_2 x, \sigma^2[1/n + x^2/n + 1/m]).$$

(D) $$\{1/\sigma^2 \mid x, y_1, v_2\} \sim \frac{1}{v}[\chi_\nu^2/\nu].$$

(E) $$\{\bar{y}_2 \mid x, y_1, v_2\} \sim t_\nu(\hat{\beta}_1 + \hat{\beta}_2 x, v[1/n + x^2/n + 1/m]).$$

We now prove these facts (if $m = 1$, the following proofs are made valid by setting $v_2 = 0$).

(A) The likelihood function associated with (1.1) is

$$\ell(\beta, \sigma^2, x \mid y_1, y_2) \propto \ell(\beta, \sigma^2 \mid y_1)\ell(\beta, \sigma^2, x \mid y_2)$$
$$\propto \ell(\beta, \sigma^2 \mid y_1)\frac{1}{\sigma^m}\exp\{-[\nu_2 v_2 + m(\bar{y}_2 - \beta_1 - \beta_2 x)^2]/2\sigma^2\}$$

Hence $(y_1, v_2, \bar{y}_2)$ is a sufficient statistic and $p(x \mid y_1, y_2) = p(x \mid y_1, v_2, \bar{y}_2)$. By Bayes theorem

$$\begin{aligned} p(x \mid y_1, v_2, \bar{y}_2) &\propto p(x \mid y_1) f(v_2, \bar{y}_2 \mid x, y_1) \\ &= p(x \mid y_1) f(\bar{y}_2 \mid x, y_1, v_2) f(v_2 \mid x, y_1). \end{aligned}$$

Fact (A) follows from this because $f(v_2 \mid x, y_1)$ does not depend on x, and $p(x \mid y_1) = p(x)$.

(B) First note that conditional on (σ^2, y_1), (x, v_2) and β are independent. From this and the Bayes solution to the ordinary regression problem [6, p. 343], we can conclude that

$$\begin{aligned} \{\beta \mid \sigma^2, x, y_1, v_2\} &\sim \{\beta \mid \sigma^2, y_1\} \\ &\sim N(\hat{\beta}, \sigma^2 I/n), \end{aligned} \tag{2.5}$$

where I is the 2×2 identity matrix. Now clearly

$$\{\bar{\epsilon}_2 \mid \sigma^2, x, y_1, v_2, \beta\} \sim N(0, \sigma^2/m). \tag{2.6}$$

Fact (B) follows from (2.5) and (2.6).

(C) This follows from (B) and the fact that $\bar{y}_2 = \beta_1 + \beta_2 x + \bar{\epsilon}_2$.

(D) Let $h = 1/\sigma^2$. By Bayes theorem

$$\begin{aligned} p(h \mid x, y_1, v_2) &\propto p(h \mid x, y_1) f(v_2 \mid h, x, y_1) \\ &= p(h \mid y_1) f(v_2 \mid h). \end{aligned} \tag{2.7}$$

But from [6, p. 343]

$$p(h \mid y_1) \propto h^{\nu_1/2 - 1} e^{-\nu_1 v_1 h/2}; \tag{2.8}$$

and it is well known that as a function of h

$$f(v_2 \mid h) \propto h^{\nu_2/2} e^{-\nu_2 v_2 h/2}. \tag{2.9}$$

Fact (D) follows from (2.7), (2.8), and (2.9).

(E) This follows from (C), (D), and Lemma 1.

Now from (A) and (E) we have

$$p(x \mid y_1, y_2) \propto \frac{p(x)}{[v(1/n + x^2/n + 1/m)]^{1/2} \left\{1 + \dfrac{[\bar{y}_2 - \hat{\beta}_1 - \hat{\beta}_2 x]^2}{v v[1/n + x^2/n + 1/m]}\right\}^{(\nu+1)/2}}. \tag{2.10}$$

Some algebraic manipulation of (2.10) yields Theorem 1.

Some remarks about Theorem 1 are now in order. The function $L(x)$ is a kind of likelihood function representing the information about x obtained from all sources except the prior distribution of x. It is clear that for fixed m and n, $L(x)$ becomes sharper (more dispersed) as F gets larger (smaller). Another property of $L(x)$ is that as $|x| \to \infty$, $L(x) = 0(1/|x|)$; so

$$\int_{-\infty}^{+\infty} L(x) = +\infty. \tag{2.11}$$

Now one can choose sequences, $\{B_{1k}\}$ and $\{B_{2k}\}$ $k=1, 2, \cdots$, of positive numbers so that if, *a priori*, x is uniformly distributed on $[-B_{1k}, B_{2k}]$, then $\lim_{k\to\infty} E(x|y_1, y_2)$ is equal to anything one wants. The point is that an improper uniform prior on x leads to a nonsensical Bayes estimator. In view of (2.11), the same can be said of the Bayes interval estimator. So it seems that a proper prior for x is a prerequisite to sensible use of the Bayes solution in Theorem 1.

More insight into $L(x)$ is obtained by writing

$$L(x) = \frac{1}{[1 + n/m + x^2]^{1/2}\left[1 + \dfrac{F(x - \hat{x})^2}{\nu(1 + n/m + x^2)}\right]^{(\nu+1)/2}} \tag{2.12}$$

Now it can be shown that $L'(x)=0$ iff x is the solution to a cubic equation, which may or may not have all real roots. If they are all real, $L(x)$ has two local maxima. Careful examination of both (2.12) and (2.4) reveals that the maximum of $L(x)$ lies between $R\hat{x}$ and $\hat{x}$ (note that $|R\hat{x}-\hat{x}|\to 0$ as $F\to\infty$). The sign of the other local maximum, if it exists, will be opposite that of $\hat{x}$. The case where $L(x)$ has two local maxima is not equivalent to the case where the confidence set in (1.12) is the union of two disjoint half-lines. For example, if $m=1$, $n=9$, and $\hat{x}=1$ (see Figure 1), then $L(x)$ is unimodal for all values of F. If m, n, and F are held fixed, then $L(x)$ can be made bimodal by taking $\hat{x}$ sufficiently large.

3. CHARACTERIZATION OF THE INVERSE ESTIMATOR

In the case $m=1$ the inverse estimator can be characterized by the following theorem:

Theorem 2. If, a priori

$$\text{[illegible]} \tag{3.1}$$

then, a posteriori,

$$\{x \mid y_1, y_2\} \sim t_{n-2}\left(\hat{x}_1, \left[\frac{n + 1 + \hat{x}_1^2/R}{(F + n - 2)}\right]\right). \tag{3.2}$$

Proof. By hypothesis,

$$\begin{aligned} p(x) &\propto \frac{1}{[1 + x^2/(n + 1)]^{(n-2)/2}} \\ &\propto \frac{1}{[1 + n + x^2]^{\nu_1/2}}. \end{aligned} \tag{3.3}$$

Now from Theorem 1 we have

$$p(x \mid y_1, y_2) \propto \frac{1}{[1 + n + R\hat{x}^2 + (F/(n - 2) + 1)(x - R\hat{x})^2]^{(n-1)/2}}. \tag{3.4}$$

But

$$\begin{aligned} R &= 1/[1 + (n-2)/F] \\ &= \frac{n\hat{\beta}_2^2}{\sum_i (y_{1i} - \bar{y}_1 - \hat{\beta}_2 x_i)^2 + n\hat{\beta}_2^2} \\ &= \frac{n\hat{\beta}_2^2}{\sum_i (y_{1i} - \bar{y}_1)^2}; \end{aligned} \tag{3.5}$$

so

$$\begin{aligned} R\hat{x} &= \left[\frac{n\hat{\beta}_2^2}{\sum_i (y_{1i} - \bar{y}_1)^2}\right]\left[\frac{y_2 - \bar{y}_1}{\hat{\beta}_2}\right] \\ &= \hat{x}_I. \end{aligned} \tag{3.6}$$

The result follows from (3.6) and (3.4).

Theorem 2 is important in view of the fact that in the literature [4], $\hat{x}_I$ has been given serious consideration as an alternative to $\hat{x}$. This theorem provides a better understanding of the inverse estimator as well as Bayes estimators in general. From (3.6) we have

$$\hat{x}_I = \left[\frac{F}{F + (n-2)}\right]\hat{x}. \tag{3.7}$$

So $\hat{x}_I$ can be viewed as a shift of $\hat{x}$ toward 0 (the prior mean of x). The relative magnitude of the shift ($|\hat{x} - \hat{x}_I| / |\hat{x}| = (n-2)/[F+(n-2)]$) is a decreasing function of F. So the more informative the data (i.e., the larger F), the less the adjustment of $\hat{x}$ toward the prior mean. Theorem 2 also says that from a Bayesian point of view, $\hat{x}_I$ cannot be justified unless the prior distribution given by (3.1) is an acceptable quantification of the prior uncertainty about x. This certainly restricts the applicability of $\hat{x}_I$. If (3.1) can be justified in a particular application, then the $100(1-\alpha)\%$ SPI can be computed easily from (3.2), and is given by

$$\hat{x}_I \pm \left\{F_{\alpha;1,n-2}\left[\frac{1 + n + \hat{x}_I^2/R}{F + n - 2}\right]\right\}^{1/2} \tag{3.8}$$

A comparison of (3.8) and (1.12) is shown in Figure 1 for the case $\alpha = .05$, $n = 9$, $\hat{x} = 1$. Note that for very large F, the two intervals are about the same; but when $F \to 0$, the SPI approaches the prior 95 percent interval estimate.

4. SIMULATED NUMERICAL EXAMPLES

In this section, we analyze three sets of data which were chosen from 16 computer simulations of model (1.1) with $n = 9$, $m = 1$, $\beta_1 = 0$, $\beta_2 = 1$, $\sigma = .698$, $x = 1$, and the x_i's equally spaced. The preceding value of σ was chosen so that the probability of S (with $\alpha = .05$) having infinite length (i.e., $P\{F < F_{.05;1,7}\}$) is .1. The two prior densities of x which are used in the analyses have the form

$$p(x) \propto \begin{cases} \dfrac{1}{\left[1+\dfrac{1}{d}\left(\dfrac{x-c}{s}\right)^2\right]^{(d+1)/2}} & -5.4 \le x \le 7.4 \\ 0 \quad \text{otherwise,} \end{cases} \tag{4.1}$$

with Prior 1 and Prior 2 specified by ($c=0$, $s=10$, $d=4$) and ($c=0$, $s=1.29$, $d=6$), respectively. Note that Prior 2 is essentially the same as the prior in Theorem 2. Truncated priors were used to simplify computations.

Table 1 lists the pertinent statistics for the three simulations. Note that Simulations 1, 2, and 3 correspond to small, medium, and large (relative to $F_{.05;1,7}=5.59$) values of F. Table 2 lists various numerical descriptions of the posterior distributions generated by the three simulations and the two-priors. For comparison, the 25th percentile, median, and 75th percentile of the two priors are (−2.18, .79, 3.84) and (−.91, 0, .92) respectively.

The data, the two regression lines, and y_2, $\hat{x}$, $\hat{x}_I$ for Simulations 1 and 3 are plotted in Figures 2 and 3 respectively. Figure 4(5) shows Prior 1(2) along with the corresponding posterior densities from Simulations 1 and 3.

Table 1. STATISTICS FOR THREE SIMULATIONS

Statistic	*Simulation 1*	*Simulation 2*	*Simulation 3*
$\hat{\beta}_1$	.179	.349	−.219
$\hat{\beta}_2$	.70	1.127	.893
v	.856	.813	.193
$\hat{x}$	.821	.549	1.073
$1/\hat{\delta}$	1.817	1.673	1.060
$\hat{x}_I$	.349	.366	.903
F	5.161	14.069	37.290
y_2	.755	.968	.740
(x_L, x_U)	$(-\infty, +\infty)$	(−1.718, 3.542)	(−.147, 2.671)

Table 2. NUMERICAL DESCRIPTIONS OF POSTERIORS

Numerical description	*Simulation 1*		*Simulation 2*		*Simulation 3*	
	Prior 1	*Prior 2*	*Prior 1*	*Prior 2*	*Prior 1*	*Prior 2*
Mean	.97	.35	.67	.37	1.16	.90
Median	.86	.35	.59	.37	1.10	.90
Mode	.70	.35	.54	.37	1.05	.90
Standard deviation	2.22	1.08	1.43	.82	.80	.59
Skewness coefficient	.15	.03	.47	.01	.92	.00
25th percentile	−.29	−.41	−.07	−.13	.71	.55
75th percentile	2.16	1.00	1.31	.86	1.53	1.26
95% shortest posterior interval (SPI)	(−3.61, 6.12)	(−1.82, 2.52)	(−2.26, 3.90)	(−1.27, 2.01)	(−.30, 2.75)	(−.28, 2.08)
95% even tailed posterior interval	(−3.75, 5.99)	(−1.82, 2.52)	(−2.14, 4.04)	(−1.27, 2.01)	(−.19, 2.91)	(−.28, 2.08)
Posterior probability of 95% confidence set S	1.00	1.00	.932	.989	.947	.959

Figure 2. DATA AND REGRESSION LINES FOR SIMULATION 1

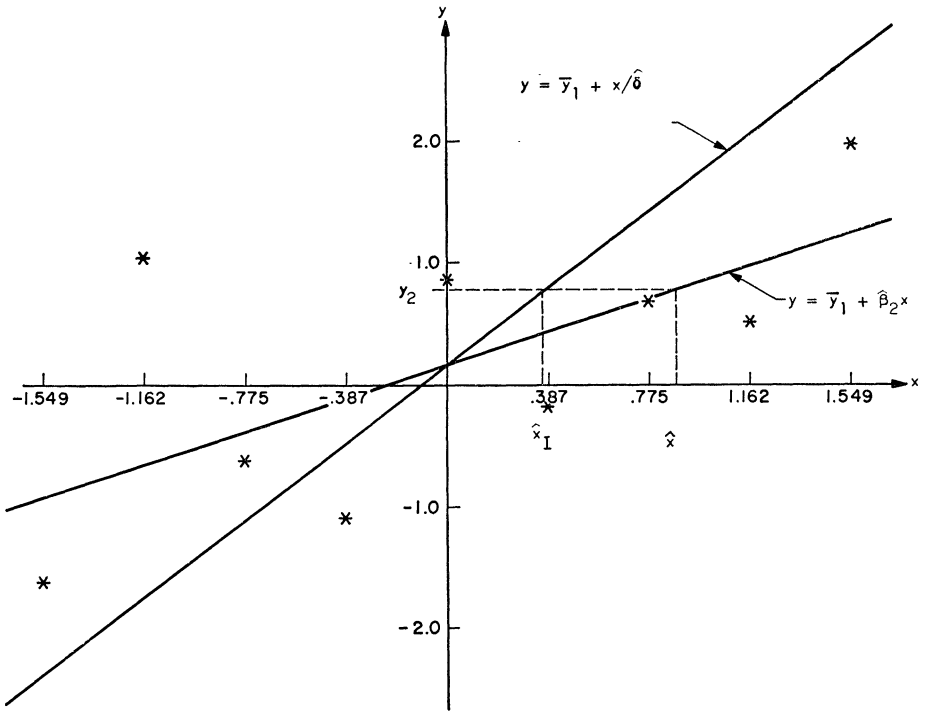


Figure 3. DATA AND REGRESSION LINES FOR SIMULATION 3

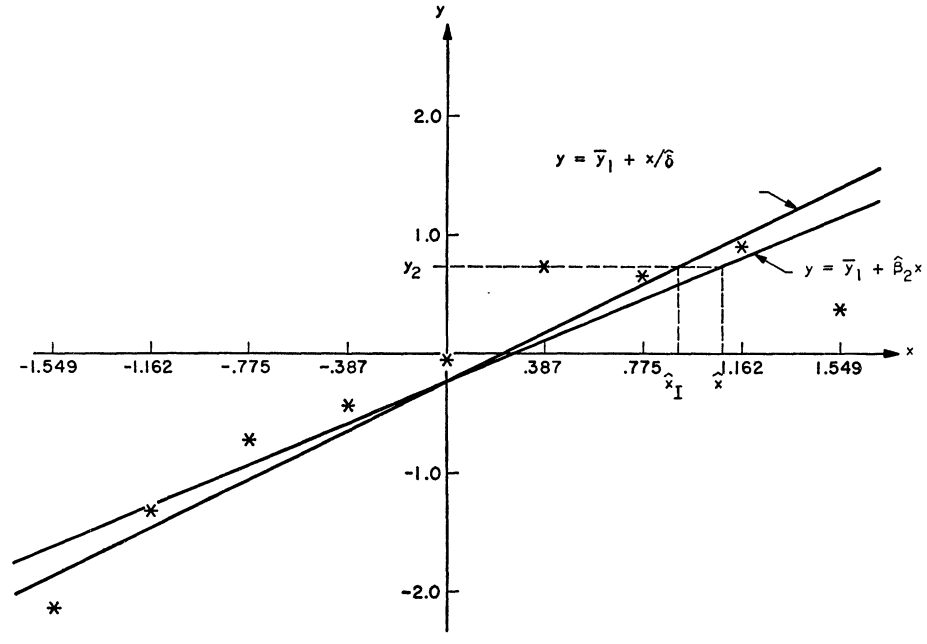


Figure 4. PRIOR 1 AND CORRESPONDING POSTERIORS FOR SIMULATIONS 1 AND 3

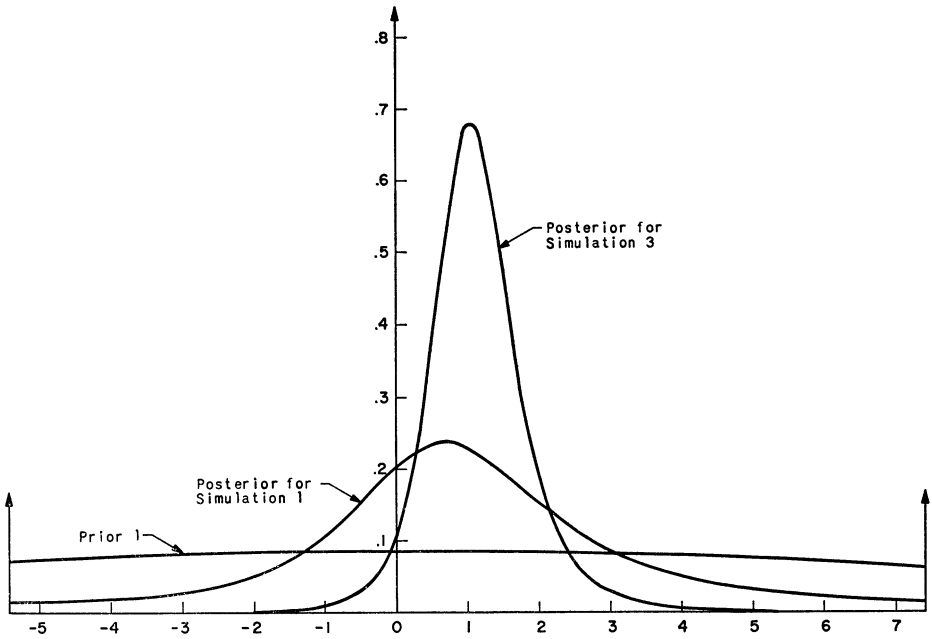
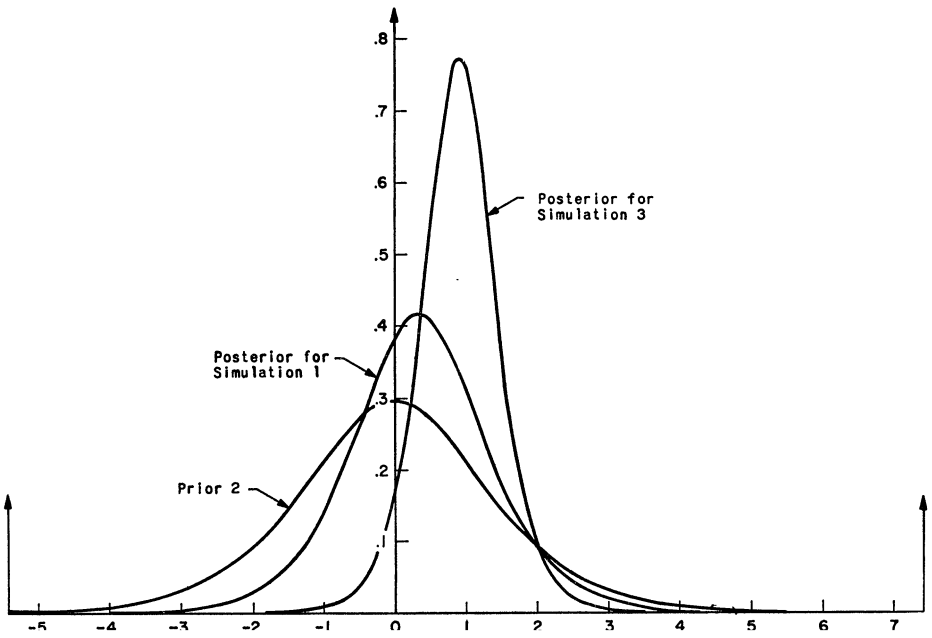


Figure 5. PRIOR 2 AND CORRESPONDING POSTERIORS FOR SIMULATIONS 1 AND 3



This analysis clearly shows that inferences about x can be made even when $F < F_{.05;1,7}$. A glance at Figure 2 suggests that these data do contain information about x, and a glance at Figure 4 confirms this feeling.

5. CONCLUSIONS

In this article, a class of solutions to the inverse linear regression problem has been presented. To select a solution, one must quantify his prior information about x. With some extra effort the model could be expanded to also allow injection of prior information about (β, σ^2).

However, the main point in the paper is that the Bayesian approach has led to valuable insight and understanding. Conditioning on the data actually observed is just the right thing to do in this problem; because if the data are weak (F small), appropriate adjustments to the MLE and classical confidence set are required, but if the data are strong (F large), not much adjustment is necessary. The prior distribution on x is acting as an anchor. It keeps the inference under control whenever the data get out of control.

In particular, it was shown that the inverse estimator, $\hat{x}_I$, is Bayes with respect to a particular informative prior on x and $|\hat{x} - \hat{x}_I| \to 0$ as $F \to \infty$.

REFERENCES

[1] Creasy, M. A., "Limits for the Ratio of Means," *Journal of the Royal Statistical Society*, Series B, 16 (1954), 186–94.

[2] Dunsmore, I. R., "A Bayesian Approach to Calibration," *Journal of the Royal Statistical Society*, Series B, 30 (1968), 396–405.

[3] Fieller, E. C., "Some Problems in Interval Estimation," *Journal of the Royal Statistical Society*, Series B, 16 (1954). 175–85.

[4] Krutchkoff, R. G., "Classical and Inverse Regression Methods of Calibration," *Technometrics*, 9 (1967), 425–39.

[5] Mood, A. M. and Graybill, F. A., *Introduction to the Theory of Statistics*, 2nd ed., San Francisco: McGraw-Hill Book Co., 1963.

[6] Raiffa, H. and Schlaifer, R., *Applied Statistical Decision Theory*, Boston: Harvard Business School, 1961.

[7] Savage, L. J., *The Foundation of Statistical Inference*, London: Methuen, 1962.

[8] Williams, E. J., "A Note on Regression Methods in Calibration," *Technometrics*, 11 (1969), 189–92.