KfK 3007
Oktober 1980

Approaches to Inverse Linear Regression

R. Avenhaus, E. Höpfinger, W. S. Jewell
Institut für Datenverarbeitung in der Technik
Projekt Spaltstoffflußkontrolle

Kernforschungszentrum Karlsruhe

KERNFORSCHUNGSZENTRUM KARLSRUHE

Institut für Datenverarbeitung in der Technik
Projekt Spaltstoffflußkontrolle

KfK 3007

APPROACHES TO INVERSE LINEAR REGRESSION

by

R. Avenhaus, E. Höpfinger*) and W.S. Jewell**)

*) Kernforschungsanlage Jülich, STE, Jülich
**) University of California, Berkeley, U.S.A.

Kernforschungszentrum Karlsruhe GmbH, Karlsruhe

Kernforschungszentrum Karlsruhe GmbH
ISSN 0303-4003

ABSTRACT

Many measurement problems can be formulated as follows: First, a certain linear relationship between two variables is to be estimated by using pairs of input and output data; thereafter, the value of an unknown input variable is to be estimated given an observation of the corresponding output variable. This problem is often referred to as inverse regression or discrimination.

In this paper first non-Bayesian approaches to the problem, thereafter the Bayesian approach by Hoadley are presented. Third, a Bayesian approach by Avenhaus and Jewell is discussed which uses the ideas of credibility theory. Finally, a new Bayesian approach is presented. The advantages and disadvantages of the various approaches are put together.

ANSÄTZE ZUR INVERSEN LINEAREN REGRESSION

ZUSAMMENFASSUNG

Viele Meßprobleme können in folgender Weise formuliert werden: Es ist zuerst ein linearer Zusammenhang zwischen zwei Variablen mit Hilfe von Paaren von Eingangs- und Ausgangsdaten zu schätzen, anschließend ist der Wert einer unbekannten Eingangsvariablen zu schätzen, die zu einer beobachteten Ausgangsvariablen gehört. Probleme dieser Art werden oft als "Inverse Regressionsprobleme" oder als "Diskriminierungsprobleme" bezeichnet.

In dieser Arbeit werden zu Beginn Ansätze zur Lösung dieser Probleme ohne Verwendung der Bayes'schen Theorie dargestellt. Im Anschluß wird der Bayes'sche Ansatz von Hoadley diskutiert. Weiter wird ein Bayes'scher Ansatz von Avenhaus und Jewell dargestellt, der die Methoden der Credibility-Theorie verwendet. Schließlich wird ein neuer Bayes'scher Ansatz präsentiert. Es werden die Vorzüge und Nachteile der verschiedenen Ansätze zusammengestellt.

C O N T E N T S

FORMULATION OF THE PROBLEM

The relationship between an independent variable x and a response variable y can often be described by the linear regression model

$$y_i = \alpha+\beta\cdot x_i+\sigma\cdot u_i \ , \quad i=1,\dots,n \ ,$$

where the u_i are independently and identically distributed random variables with means zero and variances one. Usually the u_i are assumed to be normally distributed, i.e.,

$$p(u_i \leq t) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{t} \exp(-\frac{t'^2}{2})dt' \ , \quad i=1,\dots,n \ .$$

The problem is to estimate the unknown parameters α, β and σ.

The inverse linear regression problem is an extension of the above: here, in addition to the responses corresponding to the n known independent x_i, there are m further responses $z_1,\dots,z_m$, corresponding to a single unknown x. The model is

$$y_i = \alpha+\beta\cdot x_i+\sigma\cdot u_i \qquad i=1,\dots,n$$

$$z_i = \alpha+\beta\cdot x \ +\tau\cdot v_j \qquad j=1,\dots,m$$

where u_i and v_j are independently and identically distributed random variables with means zero and variances one. The problem is to make inferences about x.

Four examples of this class of problem are given below.

Calibration and Indirect Measurement of Nuclear Materials

Nuclear materials, e.g. plutonium, are extremely difficult to measure directly by chemical means. Therefore, one uses indirect methods, based upon the heat production or the number of neutrons emitted, in order to estimate the amount of material present. From well-known physical laws, we have a general relationship between these variables, but any measurement instrument based on these principles needs first to be calibrated. Usually, this calibration can be done with the aid of standard inputs, containing known amounts of nuclear materials. However, these inputs (x_i) are not generally under our control, and in some cases, may have residual imprecisions in their values.

Measurement instruments often have longer-term drifts, during which they tend to loose their original calibration. For this reason, measurement of a given production run often consists of two distinct phases: (re)calibration of the instrument, and actual indirect measurement. With a fixed amount of time available, it is of interest to determine how much time should be spent on the two phases, assuming that additional time spent on each observation reduces observational error.

Estimation of Family Incomes by Polling

We wish to estimate, through a public opinion poll, the distribution of family incomes in a certain city district. As the major part of the population will not be willing to divulge their incomes, or will give only a very imprecise figure, we look for a dependent variable which can be more easily determined. According to the literature (see, e.g. Muth (1960)), housing expenses are strongly related to family income, and, furthermore, it may be assumed that the population is less reluctant to divulge this figure, even though they may not be able to do so precisely. Clearly, to determine this relationship exactly, we must have some families in this district who are willing to give both their total income and their household expenses. On the other hand, we have strong prior information on this relationship from similar surveys, and may have general information on income distribution from census and other sources.

Missing Variables in Bayesian Regression

In a paper with this title, Press and Scott (1974) consider a simple linear regression problem in which certain of the independent variables, x_i, are assumed to be missing in a nonsystematic way from the data pairs (x_i,y_i). Then under special assumptions about the error and prior distributions, they show that an optimal procedure for estimating the linear parameters is to first estimate the missing x_i from an inverse regression based only on the complete data pairs.

Bioassay

Using the methods of bioassay the effect of substances given in several dosages on organisms is investigated. A problem of inverse regression arises if first parameters of dosage response curves have to be estimated by evaluation of observations and later on an indirect assay is carried out to determine the dosage necessary for interesting effect (see, e.g., Rasch, Enderlein, Herrendörfer (1973)).

Problems of this kind are described in textbooks on the theory of measurements and are sometimes called discrimination problems (Brownlee (1965), Miller (1966)). They differ from the subject of 'Stochastic Approximation' (see, e.g., Wasan (1969)) in that the regression function is assumed to be linear. Stochastic approximation only requires some monotony, but this advantage is invalidated by the superiority of standard methods to the stochastic approximation method in the case of linear regression functions. Therefore the procedures of stochastic approximation will not be examined in this report.

In the following first the non-Bayesian approaches to the inverse linear regression problem are presented, especially the difficulty of the infinite variances of all the estimates is outlined. Thereafter, the Bayesian approach by Avenhaus and Jewell (1975) is discussed which uses the ideas of credibility theory and which has been written down so far only in form of an internal report. Finally, a new Bayesian approach is presented here for the first time. In the conclusion the advantages and disadvantages of the various approaches are put together. The present situation may be characterized in such a way that there are promising attempts but that there is not yet a satisfying solution to the admittedly difficult problem.

NON-BAYESIAN APPROACHES

A well-known approach is first to estimate α and β. The maximum likelihood and least squares estimates of α and β based on $y_1,\ldots,y_n$ are

$$\hat{\beta} = \frac{\sum_{i=1}^{n} (x_i-\bar{x})\cdot(y_i-\bar{y})}{\sum_{i=1}^{n} (x_i-\bar{x})^2}$$

$$\hat{\alpha} = \bar{y} - \hat{\beta}\cdot\bar{x} \quad ,$$

where $\bar{y}$ and $\bar{x}$ denote the mean values of $y_1,\ldots,y_n$ and of $x_1,\ldots,x_n$ respectively. This leads to the 'classical' estimator

$$\hat{x}_C = \frac{\bar{z}-\hat{\alpha}}{\hat{\beta}} \quad .$$

It can be seen immediately that $\hat{x}_C$ coincides with the maximum likelihood estimator for $\sigma^2>0$, $\tau^2>0$ and normally distributed u_i, v_j: The likelihood function of α, β, σ, τ and x is given by

$$L(y_1,\ldots,y_n,z_1,\ldots,z_m|\alpha,\beta,\sigma,\tau,x) =$$

$$= (2\pi\sigma^2)^{-\frac{n}{2}}\cdot\exp(-\frac{1}{2\sigma^2}\cdot\sum_{i=1}^{n}(y_i-\alpha-\beta\cdot x_i)^2) \quad .$$

$$\cdot\ (2\pi\tau^2)^{-\frac{m}{2}}\cdot\exp(-\frac{1}{2\tau^2}\cdot\sum_{j=1}^{m}(z_j-\alpha-\beta\cdot x)^2) \quad .$$

The partial derivatives $\frac{\partial L}{\partial\alpha}$, $\frac{\partial L}{\alpha\beta}$ and $\frac{\partial L}{\partial x}$ assumed to equal zero yield the equations

$$-\frac{1}{\sigma^2}\cdot\sum_i(y_i-\alpha-\beta\cdot x_i) - \frac{1}{\tau^2}\cdot\sum_j(z_j-\alpha-\beta\cdot x) = 0$$

$$-\frac{1}{\sigma^2}\cdot\sum_i(y_i-\alpha-\beta\cdot x_i)\cdot x_i - \frac{1}{\tau^2}\cdot\sum_j(z_j-\alpha-\beta\cdot x)\cdot x = 0$$

$$-\frac{1}{\tau^2}\cdot\sum_j(z_j-\alpha-\beta\cdot x)\cdot\beta = 0 \quad .$$

By exclusion of $\beta=0$ one obtains $\sum(z_j-\alpha-\beta\cdot x)=0$. Hence the first two equations reduce to the usual equations for the last square estimators $\hat{\alpha}$ and $\hat{\beta}$. The solution of the third equation is then given by $\hat{x}_C$.

One cannot judge this 'classical' criterion of minimizing the mean-square deviations, however, because of

$$E((\hat{x}_C-x)^2|\alpha,\beta,\sigma,\tau,x) = +\infty ,$$

and furthermore, because $\hat{x}_C$ has an undefined expectation value.

Krutchkoff (1967) proposed the inverse estimator $\hat{x}_I$ defined by

$$\hat{x}_I = \hat{\gamma} + \hat{\delta}\cdot\bar{z} ,$$

where

$$\hat{\delta} = \frac{\sum\limits_i (y_i-\bar{y})\cdot(x_i-\bar{x})}{\sum\limits_i (y_i-\bar{y})^2}$$

$$\hat{\gamma} = \bar{x} - \hat{\delta}\cdot\bar{y}$$

are the least squares estimators of the slope and intercept when the x_i's are formally regressed on the y_i's. Although the mean square error of $\hat{x}_I$ is finite, Williams (1969) doubted the relevance of $\hat{x}_I$. He showed that if $\sigma^2(=\tau^2)$ and the sign of β are known, then the unique unbiased estimator of x has an infinite variance. This result led him to the conclusion that since any estimator that could be derived in a theoretically justifiable manner would have an infinite variance the fact that Kruttchkoff's estimator had a finite variance seemed to be of little account.

Williams suggested to use confidence limits which should provide what is required for inverse linear regression. Hence the two papers of Perng and Tong (1974 and 1977) could meet his approval. They treated the problem of the allocation of n and m for the interval estimation of x so that the probability of coverage is maximized when the total number of observations n+m is fixed and is large.

An independent discussion of the inverse linear regression problem was given by Hoadley (1970) for $\sigma=\tau$. Part of his results will be presented in the following.

Without loss of generality it is assumed

$$\sum_i x_i = 0 \ .$$

The maximum likelihood estimators of σ based on y alone, z alone and both y and z are

$$v_1 = \frac{1}{n-2}\cdot\sum_i (y_i-\hat{\alpha}-\hat{\beta}\cdot x_i)^2$$

$$v_2 = \frac{1}{m-1}\cdot\sum_j (z_j-\bar{z})^2$$

$$v = \frac{1}{n-2+m-1}\cdot[(n-2)\cdot v_1+(m-1)\cdot v_2] \ .$$

The F-statistic, defined by

$$F := \frac{n\cdot\hat{\beta}^2}{v}$$

where $\hat{\beta}$ is the maximum likelihood estimator of β, is often used for testing the hypothesis $\beta=0$, as in fact under this hypothesis F is F-distributed with 1 and n+m degrees of freedom.

In case of m=1 a confidence set S is derived form the fact that

$$\hat{\beta}\cdot(\hat{x}_C-x)\cdot\sqrt{\frac{n}{v\cdot(n+1+x^2)}}$$

has a t-distribution with n-2 degrees of freedom. If $F_{\alpha;1,\nu}$ is the upper α point of the F-distribution with 1 and ν degrees of freedom, one gets

$$S = \begin{cases} \{x:\ x_L\le x\le x_U\} & \text{if } F>F_{\alpha;1,n-2} \\ \{x:\ x\le x_L\}\cup\{x:\ x\ge x_U\} & \text{if } \dfrac{n+1}{n+1+\hat{x}_C^2}\cdot F_{\alpha;1,n-2}\le F<F_{\alpha;1,n-2} \\ (-\infty,\ +\infty) & \text{if } F<\dfrac{n+1}{n+1+\hat{x}_C^2}\cdot F_{\alpha;1,n-2} \end{cases} ,$$

where x_L and x_U are equal to

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

with $x_L < x_U$. A graphical display of S is given in Figure 1 for n=9, $\hat{x}=1$.

As we see, this confidence set is not very helpful if

$$F < \frac{n+1}{n+1+\hat{x}_C^2}\cdot F_{\alpha;1,n-2} .$$

In this case $\hat{\beta}$ is not significantly different from zero, which may tempt one to conclude that the data provide no information about x.

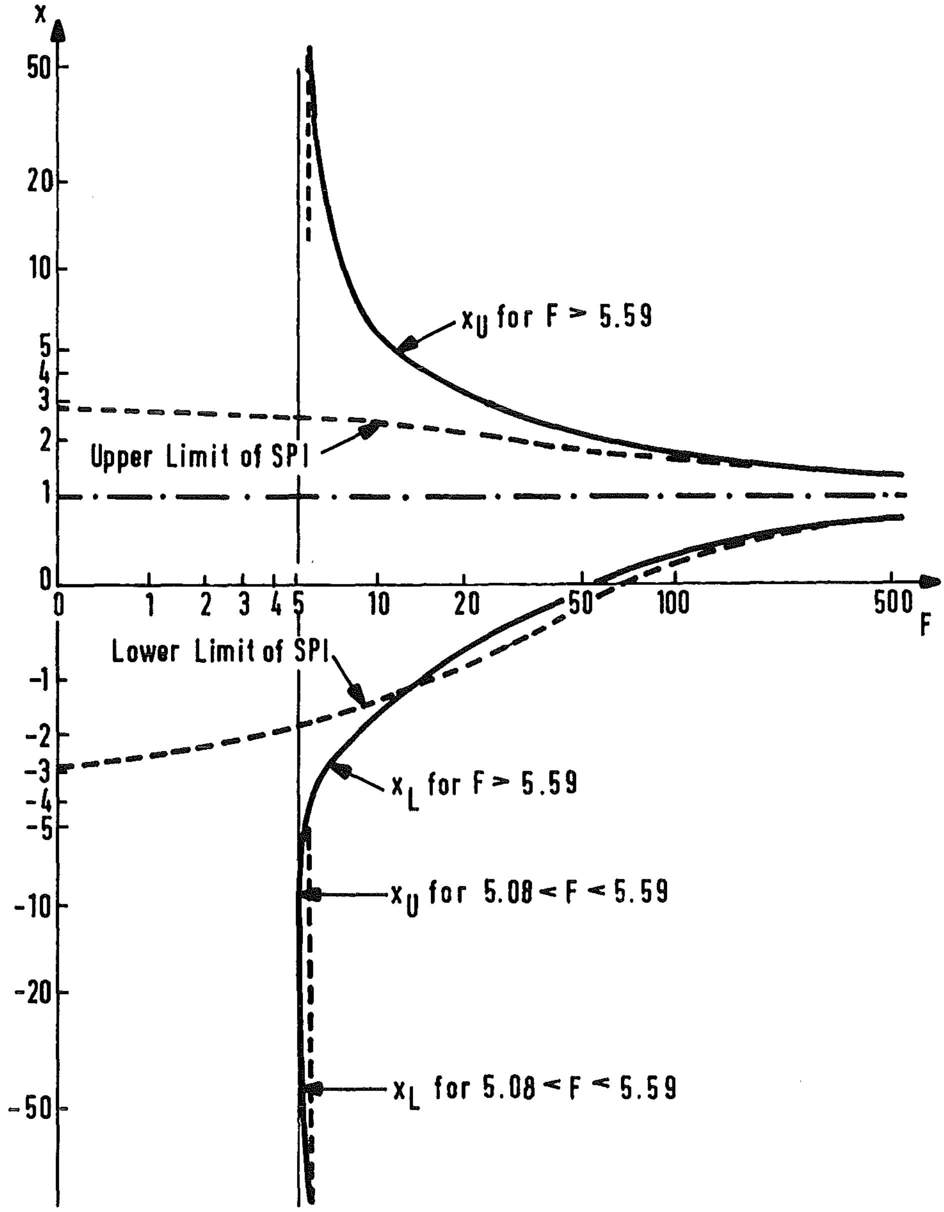

Figure 1: Comparison of 95 % Confidence Set and 95 % Shortest Posterior Interval (SPI) for n=9, $\hat{x}$=1 (after Hoadley (1970)).

THE BAYESIAN APPROACH BY HOADLEY

The situation is changed substantially if Bayesian rules are admitted. Since Bayesian rules are usually biased, the absence of an unbiased estimator for x with finite variance does not matter. Furthermore, whenever $F>0$ a shortest posterior interval can be obtained from the posterior distribution of x after the observation of $y_1 \ldots y_n$ and $z_1 \ldots z_n$.

For the sake of completeness some properties of Bayesian rules will be derived as given, e.g., by Ferguson (1967). Let $\theta \in \Theta$ denote the state chosen by nature. Given the prior distribution ψ on Θ, we want to choose a nonrandomized decision rule d that minimizes the Bayesian risk

$$r(\psi,d) := \int[\int \ell(\theta,d(k))dF_K(k|\theta)]d\psi(\theta) \quad ,$$

where $\ell(.)$ denotes the loss function and $F_K(.|\theta)$ the distribution function conditional on the chosen θ. A choice of θ by the distribution ψ, followed by a choice of the observation K from the distribution $F_K(.|\theta)$ determines in general a joint distribution of θ and K, which in turn can be determined in general by first choosing K according to its marginal distribution

$$F_K(k) = \int F_K(k|\theta)d\psi(\theta)$$

and then choosing θ according to the conditional distribution of θ, given $K=k$, $\psi(.|k)$. Hence by a change in the order of integration we may write

$$r(\psi,d) = \int[\int \ell(\theta,d(k))d\psi(\theta|k)]dF_K(k) \quad .$$

Given that these operations are admitted, it is easy now to describe a Bayesian decision rule. To find a function d(.) that minimizes the last double integral, we may minimize the inside integral separately for each k; that is, we may find for each k the decision, call it d(k), that minimizes

$$\int \ell(\theta,d(k))d\psi(\theta|k) \quad ,$$

i.e., the Bayesian decision rule minimizes the posterior conditional expacted loss, given the observation.

In the case of the inverse linear regression problem let $p(\theta)$ and $p(\theta|\text{data})$ denote the prior and posterior density of the unknown parameter θ, respectively. It is assumed that $(\alpha,\beta,\ell n\ \sigma)$ has a uniform distribution, i.e.,

$$p(\alpha,\beta,\sigma^2) \propto \frac{1}{\sigma^2} \quad . \quad ^{*)}$$

The most important results of Hoadley are given in form of the following two Theorems.

Theorem 1

Suppose that, a priori, x is independent of (α,β,σ^2), and that the prior distribution of (α,β,σ^2) is specified by

$$p(\alpha,\beta,\sigma^2) \propto \frac{1}{\sigma^2} \quad .$$

Then the posterior density of x is given by

$$p(x|y_1,\ldots,y_n,\ z_1,\ldots,z_m) \propto p(x)\cdot L(x) \quad ,$$

where

$$L(x) = \frac{\left(1+\frac{n}{m}+x^2\right)^{\frac{m+n-3}{2}}}{\left(1+\frac{n}{m}+R\cdot\hat{x}_C^2+\left(\frac{F}{m+n-3}+1\right)\cdot(x-R\cdot\hat{x}_C)^2\right)^{\frac{m+n-2}{2}}} \quad ,$$

and where

$$R = \frac{F}{F+m+n-3} \quad .$$

□

The function L(.) is a kind of likelihood function representing the information about x obtained from all sources except for the prior distribution of x. As it turns out L(.) has a lot of unpleasant properties. It seems that a proper prior for x is a prerequisite to sensible use of the Bayesian solution in the preceding theorem.

In the case m=1 the inverse estimator $\hat{x}_I$ can be characterized by the following

---

*) In Bayesian inference, the notation $u \propto v$ indicates that the function u is up to a proportional factor equal to v.

Theorem 2

If, a priori,

$$x = t_{n-3} \cdot \sqrt{\frac{n+1}{n-3}} ,$$

where the random variable t_{n-3} has a t-distribution with n-3 degrees of freedom, then, a posteriori, x conditional on $y_1, \ldots, y_n$, $z_1, \ldots, z_m$ has the same distribution as

$$\hat{x}_I + t_{n-2} \cdot \sqrt{\frac{n+1+\frac{\hat{x}_I^2}{R}}{F+n-2}} ,$$

where t_{n-2} has a t-distribution with n-2 degrees of freedom.

□

This Theorem provides a better understanding of the inverse estimator $\hat{x}_I$ as well as of Bayesian estimators in general. It seems that this result has not yet been extended to a broader class of informative priors due to technical difficulties. The papers by Halperin (1970), Kalotay (1971) and Martinelle (1970) treat other aspects and do not extend the Bayesian approach.

The following two approaches start from a Bayesian point of view, too. By restriction of the class of admitted estimators they need only the knowledge of some moments instead of the whole a priori distribution of α, β, σ, τ, and x.

THE TWO-STAGE LINEAR BAYESIAN APPROACH ACCORDING TO AVENHAUS AND JEWELL (1975)

With the help of this approach the problem is solved in two stages.[1]) At the *first stage* estimators for α and β, which are linear in $y_1,\ldots,y_n$, are constructed in such a way that a quadratic loss function is minimized. At the *second stage* an estimator for x, which is linear in the observations $z_1\ldots z_m$ is constructed in such a way that a second quadratic funtional is minimized. Since the apriori expected value of the variance σ^2 is not updated, only the apriori first and second moments of α, β, σ, τ, and x are needed.

Generally the procedure may be described as follows: Let $\theta\epsilon\Theta$ denote the state chosen by nature, and let ψ denote the prior distribution on Θ. Using a quadratic loss function

$$\ell(\theta,d) = \text{const.}(\theta-d)^2$$

for the decision d, the posterior quadratic loss $E(e(\theta,d)|K=k)$ for given observation K=k is merely the second moment about d of the posterior distribution of θ given k:

$$E(e(\theta,d)|K=k) = \int\ell(\theta,d)d\psi(\theta|k) = \text{const.} \int(\theta-d)^2\cdot d\psi(\theta|k) \quad .$$

This posterior quadratic loss is minimized by taking d as the mean of the posterior distribution of θ given k. Hence the Bayesian decision rule is

$$d(k) = E(\theta|K=k) \quad .$$

This procedure now will be applied to the Bayesian version of the inverse linear regression problem which will be presented once more for the sake of clarity.

2m+2n+5 random variables

$$\alpha,\beta,\sigma,\tau,x,u_1,\ldots,u_n, \quad y_1,\ldots,y_n, \quad v_1,\ldots,v_m, \quad z_1,\ldots,z_m$$

are considered which are defined on a probability space $(\Omega,\mathscr{F},P)$. It is assumed that the random vectors $(\alpha,\beta,\sigma,\tau)$, x, $u_1,\ldots,u_n$, $v_1,\ldots,v_m$ are stochastically independent and that the following equations hold:

---

[1]) In the original paper by Avenhaus and Jewell (1975) only the case $\sigma=\tau$ and m=1 was considered.

$$y_i = \alpha + \beta\cdot x_i + \sigma\cdot u_i \ , \quad i=1,\dots,n$$

$$z_j = \alpha + \beta\cdot x \ + \tau\cdot v_j \ , \quad j=1,\dots,m \ .$$

It is assumed that the first and second moments of u_i and v_j are known:

$$E(u_i) = E(v_j) = 0, \quad E(u_i^2) = E(v_j^2) = 1; \quad i=1,\dots,n \ , \quad j=1,\dots,m \ .$$

In the model of decision theory the sample space is the (n+m)-dimensional Euclidean space; the statistician chooses a decision function d,

$$d: \ R^{m+n} \longrightarrow R \ ,$$

which gives for each observation of values of $y_1,\dots,y_n$, $z_1,\dots,z_m$ an estimate for x, in such a way that the Bayesian risk belonging to a loss functional ℓ,

$$\ell: \ R \times R \longrightarrow R$$

is to be minimized: Let $\psi_{\alpha,\beta,\sigma,\tau,x}$ be the apriori distribution of α, β, σ, τ and x, and let $P_{y_1,\dots,y_n,\, z_1,\dots,z_m}(.\,|\alpha,\beta,\sigma,\tau,x)$ be the conditional distribution of $y_1,\dots,y_n$, $z_1,\dots,z_m$ given α, β, σ, τ and x. Then the Bayesian risk, defined by

$$r(\psi,d(.)) = \int R(\alpha',\beta',\sigma',\tau',x',d(.))d\psi_{\alpha,\beta,\sigma,\tau,x}(\alpha',\beta',\sigma',\tau',x') \quad ,$$

where R(.) is defined by

$$R(\alpha',\beta',\sigma',\tau',x';d(.))$$

$$= \int \ell(x',d(s_1,\dots,s_n,t_1,\dots,t_m))dP_{y_1,\dots,y_n,z_1,\dots,z_m}(s_1,\dots,t_m|\alpha',\beta',\sigma',\tau',x') \ ,$$

is to be minimized.

It has been pointed out already that in the case of a quadratic loss function

$$\ell(x,d) = \text{const.}(x-d)^2$$

the solution of the minimization problem is

$$E(x|y_i = s_i, \quad z_j = t_j, \quad i=1,\dots,n, \quad j=1,\dots,m).$$

The first theorem of Hoadley given in the preceding chapter highlights the complexity of this conditional expectation. Therefore, Avenhaus and Jewell (1975) use at the *first stage* of their approach the following approximate estimate for x, which is extended here to arbitrary m,

$$\hat{x}_{AJ} := c_0(y_1,\ldots,y_n) + \sum_{j=1}^{m} c_j(y_1,\ldots,y_n)\cdot z_j \quad .$$

The functions

$$c_j:\ \mathbb{R}^n \dashrightarrow \mathbb{R}\ ,\quad j=0,1,\ldots,m\ ,$$

are determined in such a way that the mean square error of x, with $z_0:=1$ and using the definition of the conditional expectation given by

$$E(x-\sum_{j=0}^{m} c_j(y_j,\ldots,y_n)\cdot z_j)^2 :=$$

$$= \int E((x-\sum_{j=0}^{m} c_j(y_j,\ldots,y_n)\cdot z_j)^2|y_1=s_1,\ldots,y_n=s_n)dP_{y_1,\ldots,y_n}(s_1,\ldots,s_n)\ ,$$

is minimized. This is performed by first minimizing the conditional expectation of the mean square error, given by

$$E((x-\sum_{j=0}^{m} c_j(.)\cdot z_j)^2|y_1=s_1,\ldots,y_n=s_n)$$

$$= \int (r-\sum_{j=0}^{m} c_j(.)\cdot t_j)^2 dP_{x,z_1,\ldots,z_m}(r,t_1,\ldots,t_m|y_1=s_1,\ldots,y_n=s_n)\ .$$

Derivation with respect to the c_0, $c_1,\ldots,c_m$ gives

$$\int (r-\sum_{j=0}^{m} c_j\cdot t_j)dP = E(x)-c_0-\sum_{j=1}^{m} c_j\cdot E(z_j|y_1=s_1,\ldots,y_n=s_n)$$

$$\int (r-\sum_{j=0}^{m} c_j\cdot t_j)\cdot t_\ell dP = E(x\cdot z_\ell|y_1=s_1,\ldots,y_n=s_n)-c_0\cdot E(z_\ell|y_1=s_1,\ldots,y_n=s_n)$$

$$-\sum_{j=1}^{m} c_j\cdot E(z_j\cdot z_\ell|y_1=s_1,\ldots,y_n=s_n)\ ,\quad \ell=1,\ldots,m\quad .$$

Putting these derivations equal to zero, we obtain the following necessary and sufficient conditions for the $c_0, c_1, \ldots, c_m$:

$$c_0(s_1,\ldots,s_n) = E(x) - \sum_{j=1}^{m} c_j(s_1,\ldots,s_n)\cdot E(z_j|y_1=s_1,\ldots,y_n = s_n)$$

$$\sum_{j=1}^{n} c_j(s_1,\ldots,s_n)\cdot \mathrm{cov}(z_j,z_\ell|y_1=s_1,\ldots,y_n=s_n) = \mathrm{cov}(x,z_\ell|y_1=s_1,\ldots,y_n = s_n) ,$$

$$\ell=1,\ldots,m .$$

Actually it is not necessary to consider this system of m+1 unknown $c_0,c_1,\ldots,c_m$, since all relevant information of the sequence $z_1,\ldots,z_m$ is contained in the mean value

$$\bar{z} := \frac{1}{m}\cdot \sum_{j=1}^{m} z_j .$$

This can be proven as follows: Let $\bar{c}_j(s_1,\ldots,s_n)$, $j=0,1$, denote the minimizing coefficients for the case $m=1$. If we write $z_1=:\bar{z}$, then the $\bar{c}_j$, $j=0,1$, are given by

$$\bar{c}_0(s_1,\ldots,s_n) = E(x)-E(\bar{z}|y_1=s_1,\ldots,y_n = s_n)$$

$$\bar{c}_1(s_1,\ldots,s_n) = \frac{\mathrm{cov}(x,\bar{z}|y_1=s_1,\ldots,y_n=s_n)}{\mathrm{var}(\bar{z}|y_1=s_1,\ldots,y_n=s_n)} .$$

Now it can be verified easily that in the general case $m>1$

$$c_0(s_1,\ldots,s_n) := \bar{c}_0(s_1,\ldots,s_n)$$

$$c_j(s_1,\ldots,s_n) := \frac{1}{m}\cdot\bar{c}_1(s_1,\ldots,s_n) , \quad j=1,\ldots,m$$

solve the system of equations given above. Hence it suffices to consider $\bar{z}$, which means that the estimator can be written as

$$\hat{x}_{AJ} = c_0(y_1,\ldots,y_n) + c_1(y_1,\ldots,y_n)\cdot\bar{z} .$$

Explicitly the terms, which are contained in this solution, are given as follows (for the sake of simplicity we write $y_i=s_i$ instead of $y_1=s_1,\ldots,$ $y_n=s_n$):

$$E(\bar{z}|y_i=s_i) = E(\alpha|y_i=s_i) + E(\beta|y_i=s_i)\cdot E(x)$$

$$\mathrm{cov}(x,\bar{z}|y_i=s_i) = E(\beta|y_i=s_i)\cdot \mathrm{var}(x)$$

$$\mathrm{cov}(z_1,z_2|y_i=s_i) = \mathrm{var}(z_1|y_i=s_i)-E(\tau^2|y_i=s_i)$$

$$\begin{aligned}\mathrm{var}(z_1|y_i=s_i) = {} & \mathrm{var}(\alpha|y_i=s_i) + 2\cdot E(x)\cdot \mathrm{cov}(\alpha,\beta|y_i=s_i) \;+ \\ & + \mathrm{var}(x)\cdot((E(\beta|y_i=s_i))^2+\mathrm{var}(\beta|y_i=s_i)) \;+ \\ & + (E(x))^2\cdot \mathrm{var}(\beta|y_i=s_i)+E(\tau^2|y_i=s_i)\;.\end{aligned}$$

The remaining problem, which represents the *second stage* of this approach, is to determine the conditional expectations

$$E(\alpha|y_i=s_i),\; E(\beta|y_i=s_i),\; \mathrm{var}(\alpha|y_i=s_i),\; \mathrm{cov}(\alpha,\beta|y_i=s_i)$$

$$\mathrm{var}(\beta|y_i=s_i) \text{ and } E(\tau^2|y_i=s_i)\;.$$

Avenhaus and Jewell do not use the observations of y_i in order to get a better estimate for σ^2, instead they replace $E(\sigma^2|y_i=s_i)$ by the apriori moment $E(\sigma^2)$. All other terms are estimated by means of linear estimators for α and β,

$$\alpha_B := \alpha_0 + \sum_{i=1}^{n} \alpha_i\cdot y_i$$

$$\beta_B := \beta_0 + \sum_{i=1}^{n} \beta_i\cdot y_i \quad ,$$

in such a way that the expectation of the quadratic loss function,

$$E((\alpha-\alpha_0-\sum_i\alpha_i\cdot y_i)^2 + (\beta-\beta_0-\sum_i\beta_i\cdot y_i)^2)$$

is minimized with respect to the unknown α_0, α_i, β_0 and β_i. This leads to the following system of equations

$$\alpha_0 = E(\alpha) - \sum_{i=1}^{n} \alpha_i \cdot (E(\alpha)+E(\beta \cdot x_i))$$

$$\beta_0 = E(\beta) - \sum_{i=1}^{n} \beta_i \cdot (E(\alpha)+E(\beta \cdot x_i))$$

$$\sum_{j=1}^{n} \alpha_j \cdot \mathrm{cov}(y_i, y_j) = \mathrm{cov}(\alpha, \alpha+\beta \cdot x_i) \ , \quad i=1,\dots,n \ ,$$

$$\sum_{j=1}^{n} \beta_j \cdot \mathrm{cov}(y_i, y_j) = \mathrm{cov}(\beta, \alpha+\beta \cdot x_i) \ , \quad i=1,\dots,n \ .$$

It can be shown (Jewell 1975) that the solution can be written as

$$\begin{pmatrix} \alpha_B \\ \beta_B \end{pmatrix} = \left(I_2 - M \right) \cdot \begin{pmatrix} E(\alpha) \\ E(\beta) \end{pmatrix} + M \cdot \begin{pmatrix} \hat{\alpha} \\ \hat{\beta} \end{pmatrix} \ ,$$

where

$$I_2 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} ,$$

$$M = C \cdot x^T \cdot x \cdot [I_2 \cdot E(\sigma^2) + C \cdot x^T \cdot x]^{-1}$$

$$C = \begin{pmatrix} \mathrm{var}(\alpha) & \mathrm{cov}(\alpha, \beta) \\ \mathrm{cov}(\alpha, \beta) & \mathrm{var}(\beta) \end{pmatrix} \ ,$$

$$x = \begin{pmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_n \end{pmatrix} \ ,$$

$$\begin{pmatrix} \hat{\alpha} \\ \hat{\beta} \end{pmatrix} = (x^T \cdot x)^{-1} \cdot x^T \cdot \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix} \ .$$

Now, $E(\alpha|y_i=s_i)$ and $E(\beta|y_i=s_i)$ are estimated by $\alpha_B(s_1,\ldots,s_n)$ and $\beta_B(s_1,\ldots,s_n)$ respectively. The second moments of α and β, i.e. the covariance matrix

$$\begin{pmatrix} \mathrm{var}(\alpha|y_i=s_i) & \mathrm{cov}(\alpha,\beta|y_i=s_i) \\ \mathrm{cov}(\alpha,\beta|y_i=s_i) & \mathrm{var}(\beta|y_i=s_i) \end{pmatrix}$$

is estimated by $M\cdot(x^T\cdot x)^{-1}$.

As already mentioned, the method does not use an aposteriori estimate for σ^2. This might easily be changed if one assumes that the apriori estimate was derived from a trial with a known number N of observations $y_1,\ldots,y_N$. Also a multiple of observations of z could be used for the estimation of σ^2 in the case of $\sigma=\tau$. Furthermore the problem has to be reconsidered whether or not the loss function for the estimation of α and β is appropriate.

A QUADRATIC BAYESIAN APPROACH

This approach tries to maintain the property of linear Bayes estimators insofar as only some moments have to be known and not an apriori distribution of α, β, σ, τ and x.

The idea is the following: Instead estimating the parameters α, β and τ of the relation

$$z_j = \alpha+\beta\cdot x+\tau\cdot v_j \ , \qquad j=1\dots m \ ,$$

the parameters of the transformed relation

$$x = \gamma+\delta\cdot z_j+w_j \ , \quad j=1\dots m \ ,$$

where

$$\gamma = -\frac{\alpha}{\beta} \ , \qquad \delta = \frac{1}{\beta} \ , \qquad w_j = -\frac{\tau}{\beta}\cdot v_j \ ,$$

are estimated by estimators which are linear in y_i, $i=1\dots n$.

Explicitely the estimator for x is given by

$$\hat{x}_Q = \hat{\gamma}+\sum_{j=1}^{m}\hat{\delta}_j\cdot z_j \quad ,$$

where the estimators $\hat{\gamma}$ and $\hat{\delta}_j$,

$$\hat{\gamma} = d_{00}+\sum_{i=1}^{n} d_{i0}\cdot y_i$$

$$\hat{\delta}_j = d_{0j}+\sum_{i=1}^{n} d_{ij}\cdot y_i \ , \qquad j=1\dots m \ ,$$

are determined in such a way that the Bayes risk, belonging to the quadratic loss,

$$\int(-x+\hat{\gamma}+\sum_{j=1}^{m}\hat{\delta}_j\cdot z_j)^2 \, dP = \int(-x+\sum_{j=0}^{m}\ \sum_{i=0}^{n} d_{ij}\cdot y_i\cdot z_j)^2 \, dP \ ,$$

where $z_0=y_0=1$, is minimized. The solution yields a *quadratic estimator*

$$\hat{x}_Q = \sum_{j=0}^{m}\ \sum_{i=0}^{n} d_{ij}\cdot y_i\cdot z_j \ ,$$

the coefficients of which are the solution of the following system of equations

$$\int (-x+\sum_{j=0}^{m}\sum_{i=0}^{n} d_{ij}\cdot y_i\cdot z_j)\cdot y_k\cdot z_\ell \, dP = 0, \quad k=0,\dots,n; \quad \ell=0,\dots,m ,$$

which is obtained by differentiating the Bayes risk partially with regard to the parameters $d_{k\ell}$. In terms of the moments of y_i and z_ℓ this system of equations has the form

$$\sum_{j=0}^{m}\sum_{i=0}^{n} d_{ij}\cdot E(y_i\cdot y_k\cdot z_j\cdot z_\ell) = E(x\cdot y_k\cdot z_\ell), \quad k=0,\dots,n; \quad \ell=0,\dots,m ,$$

which means that only the first four moments are needed.

Role of Observations z_j

It seems to be plausible that each observation z_j, $j=1,\dots,m$, should have the same importance for a 'best' estimator of x. Therefore we replace z_1 in the case of m=1 by the mean value

$$\bar{z} := \frac{1}{m}\cdot\sum_{j=1}^{m} z_j$$

and ask for the risk minimizing parameters $\bar{d}_{ij}$, $i=0,\dots,n$, $j=0,1$, of the estimators

$$\bar{\gamma} := \bar{d}_{00}+\sum_{i=0}^{n}\bar{d}_{i0}\cdot y_i$$

$$\bar{\delta} := \bar{d}_{01}+\sum_{i=1}^{n}\bar{d}_{i1}\cdot y_i$$

for γ and δ of

$$x := \gamma+\delta\cdot\bar{z}+\bar{w} .$$

Since the risk is a convex and quadratic function of $\bar{d}_{ij}$ the optimal $\bar{d}_{ij}$ are completely determined as solutions of

$$\sum_{j=0}^{1}\sum_{i=0}^{n}\bar{d}_{ij}\cdot E(y_i\cdot y_k\cdot \bar{z}^{j+\ell}) = E(x\cdot y_k\cdot \bar{z}^{\ell}) , \quad k=0,\dots,n; \quad \ell=0,1 ,$$

where $\bar{z}^0 := 1$. Since

$$E(y_i \cdot y_k \cdot z_j \cdot z_\ell) = E(y_i \cdot y_k \cdot z_1^2) - \chi(j,\ell) \cdot E(y_i \cdot y_k \cdot \tau^2) \quad , \quad j,\ell = 1,\ldots,m \; .$$

where

$$\chi(j,\ell) := \begin{cases} 0 & \text{for} \quad j \neq \ell \\ 1 & \quad j = \ell \end{cases} \quad ,$$

we get

$$E(y_i \cdot y_k \cdot \bar{z}) = E(y_i \cdot y_k \cdot z_1) \; ,$$

$$E(y_i \cdot y_k \cdot \bar{z}^2) = E(y_i \cdot y_k \cdot z_1^2) - \frac{m-1}{m} \cdot E(y_i \cdot y_k \cdot \tau^2) \; ,$$

where

$$E(y_i \cdot \tau^2) = E(\alpha \cdot \tau^2) + x_i \cdot E(\beta \cdot \tau^2) \quad , \quad i=1,\ldots,m \; ,$$

$$E(y_i^2 \cdot \tau^2) = E(\alpha^2 \cdot \tau^2) + 2 \cdot x_i \cdot E(\alpha \cdot \beta \cdot \tau^2) + x_i^2 \cdot E(\beta^2 \cdot \tau^2) + E(\sigma^2 \cdot \tau^2) \quad , \quad i=1,\ldots,m \; ,$$

$$E(y_i \cdot y_k \cdot \tau^2) = E(\alpha^2 \cdot \tau^2) + (x_i + x_k) \cdot E(\alpha \cdot \beta \cdot \tau^2) + x_i \cdot x_k \cdot E(\beta^2 \cdot \tau^2) \quad , \quad i,k=1,\ldots,n \quad i \neq k \; .$$

We show that the estimator

$$\hat{x}_Q = \sum_{j=0}^{1} \sum_{i=0}^{n} \bar{d}_{ij} \cdot (\bar{z})^j$$

represents a solution of our original problem. Let

$$d_{i0} := \bar{d}_{i0}$$

$$d_{ij} := \frac{1}{m} \cdot \bar{d}_{i1} \quad , \quad j=1,\ldots,m \quad , \quad i=0,\ldots,n \; .$$

It is easily shown that these terms solve the original system of equations

$$\sum_j \sum_i d_{ij} \cdot E(y_i \cdot y_k \cdot z_j \cdot z_\ell) = E(x \cdot y_k \cdot z_\ell) \; ;$$

therefore

$$\hat{x}_Q = \sum_{j=0}^{m} \sum_{i=0}^{n} d_{ij} \cdot y_j \cdot z_j = \sum_{j=0}^{1} \sum_{i=0}^{n} \bar{d}_{ij} (\bar{z})^j$$

is the risk minimizing estimator for x.

It should be noted that the mean value estimator is not always the single solution.

Unbiasedness

The first equation ($k=\ell=0$) of the system of equations determining d_{ij},

$$\sum_{j=0}^{m} \sum_{i=0}^{n} d_{ij} \cdot E(y_i \cdot z_{ij}) = E(x) \ ,$$

shows that the estimator is *unbiased* with regard to the apriori distribution.

One would regard the estimator as trivial if $d_{00}=E(x)$ and $d_{ij}=0$ for $(i,j)\neq(0,0)$, i.e., if the estimator neglected the observations of $y_1,\dots,y_n$, $z_1,\dots,z_m$. By inspection of the equations determining the d_{ij}, this holds if and only if

$$\mathrm{var}(x) \cdot E(\alpha \cdot \beta + \beta^2 \cdot x_k) = 0 \ , \quad k=0,\dots,n \ ,$$

or equivalently if

$$\mathrm{var}(x) = 0 \quad \text{or} \quad E(\beta^2) = 0 \ .$$

Hence these cases have to be excluded.

Computational Procedure

In the following we consider only the mean value $\bar{z}$ of observations z_j, $j=1,\dots,m$. Therefore we write z instead of $\bar{z}$ for the sake of simplicity. For the same reason we write d_{ij} instead of $\bar{d}_{ij}$, $i=0,1,\dots,n$, $j=0,1$. This can be interpreted as the description of the situation where one has only one observation z_1 with the error $\frac{1}{m} \cdot \tau^2$ instead of τ^2.

In the case that the first four joint moments of $y_1,\dots,y_n$ and z can easily be obtained, another system of equations can be used for the determination of the estimator. With the definitions

$$A := \sum_{i=1}^{n} d_{i0} \cdot y_i \ , \qquad B := \sum_{i=1}^{n} d_{i1} \cdot y_i$$

the system of $2 \cdot n+2$ equations for the coefficients d_{i0}, d_{i1}, $i=0,\ldots,n$, has the following form:

$$d_{00} + E(A) + d_{01} \cdot E(z) + E(B \cdot z) = E(x)$$

$$d_{00} \cdot E(z) + E(A \cdot z) + d_{01} \cdot E(z^2) + E(B \cdot z^2) = E(x \cdot z)$$

$$d_{00} \cdot E(y_k) + E(A \cdot y_k) + d_{01} \cdot E(\bar{z} \cdot y_k) + E(B \cdot y_k \cdot \bar{z}) = E(x \cdot y_k)$$

$$d_{00} \cdot E(y_k \cdot z) + E(A \cdot y_k \cdot z) + d_{01} \cdot E(z^2 \cdot y_k) + E(B \cdot y_k \cdot z^2) = E(x \cdot y_k \cdot z), \quad k=1,\ldots,n \ .$$

Solving the first two equations for d_{00} and d_{01}, we get

$$d_{01} = \frac{1}{var(z)} \cdot [cov(x,z) - E((A+B \cdot z) \cdot (z-E(z))]$$

$$d_{00} = \frac{1}{var(z)} \cdot [cov(x \cdot z, z) - cov(x, z^2) + E((A+B \cdot z) \cdot (z \cdot E(z) - E(z^2)))] \ .$$

Inserting these formulae into the remaining equations we get with the map $f(.,.)$ defined for each pair of random variables U,V by

$$f(U,V) := var(z) \cdot E(U \cdot V) - E(V) \cdot (cov(U \cdot z, z) - cov(U, z^2)) - E(V \cdot z) \cdot cov(U,z)$$

the following system of equations for d_{i0}, d_{i1}, $i=1,\ldots,n$:

$$\sum_{i=1}^{n} d_{i0} \cdot f(y_i, y_k) + \sum_{i=1}^{n} d_{i1} \cdot f(y_i \cdot z, y_k) = f(x, y_k)$$

$$\sum_{i=1}^{n} d_{i0} \cdot f(y_i, y_k \cdot z) + \sum_{i=1}^{n} d_{i1} \cdot f(y_i \cdot z, y_k \cdot z) = f(x, y_k \cdot z), \quad k=1,\ldots,n \ .$$

Having solved these equations, we can determine d_{01} and d_{00} as follows:

$$d_{01} = \frac{1}{var(z)} \cdot [cov(x,z) - \sum_{i=1}^{n} d_{i0} \cdot cov(y_i, z) + \sum_{i=1}^{n} d_{i1} \cdot cov(y_i \cdot z, z)]$$

$$d_{00} = \frac{1}{\mathrm{var}(z)} \cdot [\mathrm{cov}(x\cdot z,z)-\mathrm{cov}(x,z^2) + \sum_{i=1}^{n} d_{i0}\cdot(\mathrm{cov}(y_i\cdot z,z)-\mathrm{cov}(y_i,z^2))$$

$$-\sum_{i=1}^{n} d_{i1}\cdot(\mathrm{cov}(y_i\cdot z^2,z)-\mathrm{cov}(y_i\cdot z,z^2))] \ ,$$

where the moments needed explicitely are given by

$E(z) = E(\alpha)+E(x)\cdot E(\beta)$

$E(y_k\cdot z) = E(\alpha^2)+[x_k+E(x)]\cdot E(\alpha\cdot\beta)+x_k\cdot E(x)\cdot E(\beta^2)$

$E(y_i\cdot y_k) = E(\alpha^2)+[x_i+x_k]\cdot E(\alpha\cdot\beta)+x_i\cdot x_k\cdot E(\beta^2)+\chi(i,k)\cdot E(\sigma^2)$

$$E(y_i\cdot y_k\cdot z) = E(\alpha^3)+[x_i+x_k+E(x)]\cdot E(\alpha^2\cdot\beta)+[x_i\cdot x_k+x_i\cdot E(x)+x_k\cdot E(x)]\cdot E(\alpha\cdot\beta^2)+$$
$$+x_i\cdot x_k\cdot E(x)\cdot E(\beta^3)+\chi(i,k)\cdot[E(\alpha\cdot\sigma^2)+E(x)\cdot E(\beta\cdot\sigma^2)]$$

$E(z^2) = E(\alpha^2)+2\cdot E(x)\cdot E(\alpha\cdot\beta)+E(x^2)\cdot E(\beta^2)+ \frac{1}{m}\cdot E(\tau^2)$

$$E(y_k\cdot z^2) = \frac{1}{m}\cdot E(\alpha\cdot\tau^2)+x_k\cdot\frac{1}{m}\cdot E(\beta\cdot\tau^2)+E(\alpha^3)+[2\cdot E(x)+x_k]\cdot E(\alpha^2\cdot\beta)+$$
$$+[2\cdot x_k\cdot E(x)+E(x^2)]\cdot E(\alpha\cdot\beta^2)+x_k\cdot E(x^2)\cdot E(\beta^3)$$

$$E(y_i\cdot y_k\cdot z^2) = E(\alpha^4)+[x_i+x_k+2\cdot E(x)]\cdot E(\alpha^3\cdot\beta)+[x_i\cdot x_k+2\cdot(x_i+x_k)\cdot E(x)+E(x^2)]\cdot$$
$$\cdot E(\alpha^2\cdot\beta^2)+[(x_i+x_k)\cdot E(x^2)+2\cdot x_i\cdot x_k\cdot E(x)]\cdot E(\alpha\cdot\beta^3)+$$
$$+x_i\cdot x_k\cdot E(x^2)\cdot E(\beta^4)+ \frac{1}{m}\cdot E(\alpha^2\cdot\tau^2)+(x_i+x_k)\cdot\frac{1}{m}\cdot E(\alpha\cdot\beta\cdot\tau^2)+$$
$$+x_i\cdot x_k\cdot\frac{1}{m}\cdot E(\beta^2\cdot\tau^2)+\chi(i,k)\cdot[E(\alpha^2\cdot\sigma^2)+E(x^2)\cdot E(\beta^2\cdot\sigma^2)+$$
$$+2\cdot E(x)\cdot E(\alpha\cdot\beta\cdot\sigma^2)+ \frac{1}{m}\cdot E(\sigma^2\cdot\tau^2)]$$

$E(x\cdot y_k) = E(x)\cdot[E(\alpha)+x_k\cdot E(\beta)]$

$E(x\cdot z) = E(x)\cdot E(\alpha)+E(x^2)\cdot E(\beta)$

$E(x\cdot y_k\cdot z) = E(x)\cdot E(\alpha^2)+[E(x^2)+x_k\cdot E(x)]\cdot E(\alpha\cdot\beta)+x_k\cdot E(x^2)\cdot E(\beta^2)$.

Special Case

Let us now assume that α and β are exactly known, i.e.,

$$E(\alpha) = \alpha, \quad E(\beta) = \beta; \quad var(\alpha) = var(\beta) = 0 \quad .$$

Then we get

$$var(z) = var(\alpha+\beta\cdot x+\tau\cdot v) = \beta^2\cdot var(x)+ \frac{1}{m}\cdot E(\tau^2)$$

$$cov(x,z) = cov(x,\alpha+\beta\cdot x) = \beta\cdot var(x)$$

$$cov(x\cdot z,z)-cov(x,z^2) = -\alpha\cdot\beta\cdot var(x)+ \frac{1}{m}\cdot E(\tau^2)\cdot E(x)$$

$$E(x\cdot y_k) = E(x)\cdot E(y_k) \quad \text{etc.,}$$

and therefore

$$f(x,y_k) = E(y_k)\cdot[E(x)\cdot(\beta^2.var(x)+ \frac{1}{m}\cdot E(\tau^2))+\alpha\cdot\beta\cdot var(x)- \frac{1}{m}\cdot E(\tau^2)\cdot E(x)$$
$$-(\alpha+\beta\cdot E(x))\cdot\beta\cdot var(x)] = 0 \quad ,$$

$$f(x,y_k\cdot z) = E(y_k)\cdot[E(\alpha\cdot x+\beta\cdot x^2)\cdot(\beta^2\cdot var(x)+ \frac{1}{m}\cdot E(\tau^2))+(\alpha+\beta\cdot E(x))\cdot(\alpha\cdot\beta\cdot var(x)$$
$$- \frac{1}{m}\cdot E(\tau^2)\cdot E(x))-(\beta^2\cdot var(x)+ \frac{1}{m}\cdot E(\tau^2)+(\alpha+\beta\cdot E(x)^2)\cdot\beta\cdot var(x)] = 0 \quad .$$

As the system of equations for the d_{i0}, d_{i1}, $i=1,\dots,n$ is homogeneous, the system has the trivial solution

$$d_{i0} = d_{i1} = 0 \quad \text{for} \quad i=1,\dots,n \quad .$$

This result is reasonable: If the parameters α and β of the regression line are exactly known, one does not need the y_i for estimating these parameters.

With A=B=0 we get

$$d_{01} = \frac{cov(x,z)}{var(z)} = \frac{\beta\cdot var(x)}{\beta^2\cdot var(z)+ \frac{1}{m}\cdot E(\tau^2)}$$

$$d_{00} = \frac{cov(x\cdot z,z)-cov(x,z^2)}{var(z)} = \frac{-\alpha\cdot\beta\cdot var(x)+ \frac{1}{m}\cdot E(\tau^2)\cdot E(x)}{\beta^2\cdot var(z)+ \frac{1}{m}\cdot E(\tau^2)} \quad .$$

Therefore x is estimated by

$$\hat{x}_Q = d_{00} + d_{01}\cdot z = \frac{-\alpha\cdot\beta\cdot\mathrm{var}(x) + \frac{1}{m}\cdot E(\tau^2)\cdot E(x) + \beta\cdot\mathrm{var}(x)\cdot z}{\beta^2\cdot\mathrm{var}(z) + \frac{1}{m}\cdot E(\tau^2)} .$$

This estimate can be written in the following intuitive form

$$\hat{x}_Q = \frac{1}{1+\frac{m\cdot\beta^2\cdot\mathrm{var}(x)}{E(\tau^2)}}\cdot E(x) + \frac{1}{1+\frac{E(\tau^2)}{m\cdot\beta^2\cdot\mathrm{var}(x)}}\cdot\frac{z-\alpha}{\beta} ,$$

which can be interpreted as follows: If the apriori information on x is much better than the measurement uncertainty, i.e., if $\frac{1}{m}\cdot E(\tau^2) \gg \beta^2\cdot\mathrm{var}(x)$, then x is simply estimated by the apriori information. In the opposite case x is estimated by inverting the regression line.

CONCLUSIONS

Four estimators of practical importance were considered, the maximum likelihood estimator $\hat{x}_C$, the inverse regression estimator $\hat{x}_I$, the two stage estimator $\hat{x}_{AJ}$, and the quadratic estimator $\hat{x}_Q$. All of them are linear in the mean value $\bar{z}$, i.e., all have a shape $\hat{x}=c_0+c_1\cdot\bar{z}$. The coefficients c_0 and c_1 depend on the observation of $y_1,\ldots,y_n$ and the apriori information. Since the expectation value of $\hat{x}_C$ does not exist, its relevance as a point estimator seems doubtful. All other estimators have their own merits and shortcomings. The estimator $\hat{x}_I$ is easily calculable but up to now justified as a Bayesian estimator only for special a-priori distribution functions. $\hat{x}_{AJ}$ uses only the first and second a-priori moments instead of the whole a-priori distribution. The numerical expenditure is substantially higher as with $\hat{x}_i$. The estimator $\hat{x}_{AJ}$ however needs further theoretical investigation. The quadratic estimator $\hat{x}_Q$ is the only one which has been derived as a solution of a risk minimizing problem. It is the only one which is linear in the observation $y_1,\ldots,y_n$. Its confidence region and sequential properties have not been investigated as yet. Furthermore even more computation effort is needed as for the other estimators. In addition the required knowledge of the third and fourth moments of the apriori distribution requires an increased effort. Whether this problem can be circumvented by a similar "semi-minimax" estimator using only the first, two apriori moments cannot be answered as yet.

So far only a few numerical calculation have been performed. They indicated that the four different methods led to not too different estimations of x however, that the coefficients c_0 and c_1 of the linear form $c_0+c_1\cdot\bar{z}$ different substantially depending on the apriori information. Thus it seems that considerable numerical work is required in order to get a feeling for the usefulness of the various approaches under given circumstances.

Contrary to the fact that already a large amount of research effort has been invested into the inverse regression problem, only a few results have been obtained especially if more general nonlinear estimators are considered. It seems that the scope of the problem of inverse linear regression has not yet been understood.

REFERENCES

Avenhaus, R., and Jewell, W.S. (1975). Bayesian Inverse Regression and Discrimination: An Application of Credibility Theory, International Institute for Applied Systems Analysis, RM-75-27, Laxenburg, Austria.

Baur, F. (1977) Universität Augsburg, unveröffentlicht.

Brownlee, K.A. (1965). Statistical Theory and Methodology in Science and Engineering, Wiley, New York.

Ferguson, Th. (1967). Mathematical Statistics - A Decision Theoretic Approach. Academic Press, New York and London.

Halperin, M. (1970). On Inverse Estimation in Linear Regression. Technometrics, Vol. 12, No. 4, pp. 727-736.

Hoadley, B. (1970). A Bayesian Look at Inverse Linear Regression. Journ. Amer. Statist. Association, Vol. 65, No. 329, pp. 356-369.

Jewell, W.S. (1975). Bayesian Regression and Credibility Theory, International Institute for Applied Systems Analysis, RM-75-63, Laxenburg, Austria.

Kalotay, A.J. (1971). Structural Solution to the Linear Calibration Problem. Technometrics, Vol. 13, No. 4, pp. 761-769.

Krutchkoff, R.G. (1967). Classical and Inverse Regression Methods of Calibration. Technometrics, Vol. 9, No. 3, pp. 425-435.

Krutchkoff, R.G. (1969). Classical and Inverse Regression Methods of Calibration in Extrapolation. Technometrics, Vol. 11, No. 3, pp. 605-608.

Martinelle, S. (1970). On the Choice of Regression in Linear Calibration. Comments on a paper by R.G. Krutchkoff, Technometrics, Vol. 12, No. 1, pp. 157-161.

Miller, R.G. (1966). Simultaneous Statistical Inference, McGraw Hill, New York.

Muth, R.F. (1960). The Demand for Non-Farm Housing, Included in: A.C. Harberger (ed.), The Demand for Durable Goods. The University of Chicago.

Perng, S.K., and Tong, Y.L. (1974). A Sequential Solution to the Inverse Linear Regression Problem. Ann. of Statistics, Vol. 2, No. 3, pp. 535-539.

Perng, S.K., and Tong, Y.L. (1977). Optimal Allocation of Observations in Inverse Linear Regression. Ann. of Statistics, Vol. 5, No. 1, pp. 191-196.

Press, S.J., and Scott, A. (1975). Missing Variables in Bayesian Regression Included in: S.E. Fienberg and A. Zellner (eds.), Studies in Bayesian Econometrics and Statistics, North-Holland, Amsterdam.

Rasch, D., Enderlein, G., Herrendörfer, G. (1973). Biometrie. Deutscher Landwirtschaftsverlag, Berlin.

Wasan, M.T. (1969). Stochastic Approximation, Cambridge at the University Press.