

AD-A017 239

CREDIBILITY REGRESSION WITH SIMPLE TRENDS

William S. Jewell

California University

Prepared for:

Army Research Office - Durham

September 1975

DISTRIBUTED BY:

NTIS

National Technical Information Service
U. S. DEPARTMENT OF COMMERCE

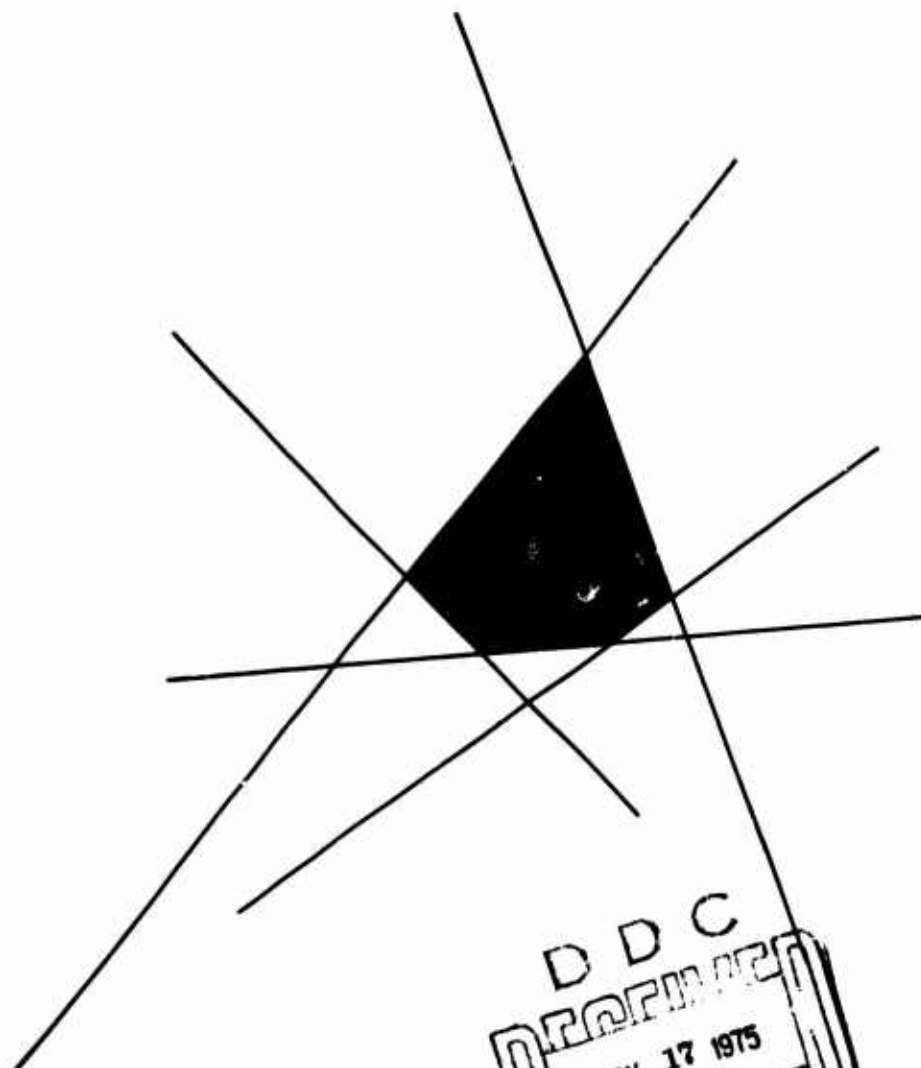
323172

CRC 75-16
SEPTEMBER 1975

CREDIBILITY REGRESSION WITH SIMPLE TRENDS

by
WILLIAM S. JEWELL

ADA 017239



D D C
 REPRODUCED
 NOV 17 1975
 REGISTERED
 C

OPERATIONS RESEARCH CENTER

Reproduced by
NATIONAL TECHNICAL
INFORMATION SERVICE
U.S. Department of Commerce
Springfield VA 22151

DISTRIBUTION STATEMENT A
 Approved for public release
 Distribution Unlimited

UNIVERSITY OF CALIFORNIA • BERKELEY

CREDIBILITY REGRESSION WITH SIMPLE TRENDS

Operations Research Center Research Report No. 75-16

William S. Jewell

September 1975

U. S. Army Research Office - Durham

DAHC04-75-G-0163

**Department of Industrial Engineering
and Operations Research
University of California, Berkeley**

**APPROVED FOR PUBLIC RELEASE;
DISTRIBUTION UNLIMITED.**

Unclassified

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER ORC 75-16	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) CREDIBILITY REGRESSION WITH SIMPLE TRENDS		5. TYPE OF REPORT & PERIOD COVERED Research Report
		6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s) William S. Jewell		8. CONTRACT OR GRANT NUMBER(=) DAHC04-75-G-0163
9. PERFORMING ORGANIZATION NAME AND ADDRESS Operations Research Center University of California Berkeley, California 94720		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS P-12549-M
11. CONTROLLING OFFICE NAME AND ADDRESS U. S. Army Research Office - Durham Box CM, Duke Station Durham, North Carolina 27706		12. REPORT DATE September 1975
		13. NUMBER OF PAGES 15
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office)		15. SECURITY CLASS. (of this report) Unclassified
		15a. DECLASSIFICATION DOWNGRADING SCHEDULE
16. DISTRIBUTION STATEMENT (of this Report) Approved for public release; distribution unlimited.		
17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)		
18. SUPPLEMENTARY NOTES		
19. KEY WORDS (Continue on reverse side if necessary and identify by block number) Regression Bayesian Analysis Credibility Theory Trend Analysis		
20. ABSTRACT (Continue on reverse side if necessary and identify by block number) (SEE ABSTRACT)		

DD FORM 1473
1 JAN 73

EDITION OF 1 NOV 65 IS OBSOLETE
S/N 0102-014-6601

//

Unclassified

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

THE FINDINGS IN THIS REPORT ARE NOT TO BE
CONSTRUED AS AN OFFICIAL DEPARTMENT OF
THE ARMY POSITION, UNLESS SO DESIGNATED
BY OTHER AUTHORIZED DOCUMENTS.

ABSTRACT

Bayesian linear regression models, where the parameters follow simple trends, can be efficiently solved using credibility approximations and recursive calculations which exploit the special structure of the problem.

1.

CREDIBILITY REGRESSION WITH SIMPLE TRENDS

by

William S. Jewell

1. BAYESIAN REGRESSION

Consider a *linear (regression) model*

$$(1.1) \quad \tilde{y} = X\tilde{\beta} + \tilde{u},$$

where \tilde{y} and \tilde{u} are $n \times 1$ random vectors of observable *output variables* and unobservable *error variables*, respectively, X is a known $n \times k$ *design matrix*, and $\tilde{\beta}$ is $k \times 1$ random vector of unknown *regression parameters*. We assume that a prior joint density of $(\tilde{\beta}, \tilde{u})$ is known. Given an observation $\tilde{y} = y$, the problem is to draw posterior-to-data inferences about $\tilde{\beta}$, or about future values of \tilde{y} for some different design matrix; this is a problem in *Bayesian regression* [see, e.g. Zellner (1971)].

Many insurance rate-making models are of this type. With one parameter (an unknown mean), and $X = [1, 1, \dots, 1]'$, we have classical credibility theory; multi-dimensional credibility theory has several such means, and X contains blocks of unit matrices. In additive relativity premium models, there are groups of different factors, with one parameter from each group added to make up the total premium; X consists of echelon patterns of 1's and 0's [Grimes (1971)].

Of particular interest in these changing times are linear models in which the parameters are subject to inflation by unknown amounts. Although there is no formal difficulty in including trends in the Bayesian framework, there are important practical difficulties, due to the necessity of providing full-information priors, and the resulting large dimensionality of real problems. This paper explores the case where the parameters follow a simple but unknown trend, and simplifications are possible by using credibility theory and an iterative computational scheme.

2. CREDIBILITY REGRESSION

A complete Bayesian regression analysis is very difficult, usually requiring restrictive distributional assumptions or complicated algebraic manipulations; [see e.g. Box and Tiao (1973), Morales (1971), and Zellner (1971).] However, recent work [Hachemeister (1974); Taylor (1974); Jewell (1975)] has shown that the linearized approach of credibility theory can be very useful for general models like (1.1), if the goal is to update only mean values; preposterior (average before-the-observation) covariances can also be determined.

Let the prior knowledge of $(\tilde{\beta}, \tilde{u})$ be summarized in the mean vectors:

$$(2.1) \quad E\{\tilde{\beta}\} = b \quad ; \quad E\{\tilde{u} \mid \tilde{\beta}\} = 0 \quad (\text{for all } \tilde{\beta}) \quad ;$$

and the covariance matrices:

$$(2.2) \quad V\{\tilde{\beta}\} = \Delta \quad ; \quad EV\{\tilde{y} \mid \tilde{\beta}\} = V\{\tilde{u}\} = E \quad ;$$

of order $k \times k$ and $n \times n$, respectively. We define also alternate-dimension versions of the covariances:

$$(2.3) \quad D = X \Delta X \quad ; \quad \epsilon = (X'E^{-1}X)^{-1} \quad ;$$

which are $n \times n$ and $k \times k$, respectively. Even if E is positive definite (most applications have E diagonal), ϵ may not exist in many linear models of interest because X is not of rank $\max(k, n)$.

Jewell (1975) shows that there are two versions of the updated credibility forecast of the mean parameter values $f(y) \approx E\{\beta | y\}$. In the first version:

$$(2.4) \quad f(y) = (I_k - ZX)b + Zy,$$

where I_k is $k \times k$ unit matrix, and Z is a $k \times n$ *credibility matrix*

$$(2.5) \quad Z = \Delta X'(E + D)^{-1}.$$

This clearly exists if, say, E is positive definite, and X contains only nonnegative elements; an $n \times n$ inversion is required, even if E^{-1} is known, hence this form is suitable for limited-observation experiments where $n < k$.

(Parenthetically, note that n refers to different *dimensions* of observations, not the actual volume of observations; if we have v_i samples, $y_{i1}, y_{i2}, \dots, y_{iv_i}$, in dimension i , we can aggregate, using $y_i = \sum_j y_{ij}/v_i$, making appropriate adjustments in E .)

In the second version, we obtain

$$(2.6) \quad f(y) = (I_k - z)b + z\hat{\beta}(y),$$

where $\hat{\beta}(y)$ is the classical (generalized) least-squares estimator of $\tilde{\beta}$:

$$(2.7) \quad \hat{\beta}(y) = \epsilon X'E^{-1}y = (X'E^{-1}X)^{-1}X'E^{-1}y,$$

and z is a $k \times k$ credibility matrix:

$$(2.8) \quad z = \left(I_k + \epsilon \Delta^{-1} \right)^{-1} = \Delta (I_k + \epsilon^{-1} \Delta)^{-1} \epsilon^{-1}.$$

This matrix is analogous to the usual multidimensional credibility matrix with "one" sample [Jewell (1974)] and gives a more readily interpreted mixing of prior mean

and classical estimator. Moreover, in many applications $k < n$, and the second form in (2.8) shows that only one $k \times k$ inversion is required to find $f(y)$, if E^{-1} is known, which greatly reduces the computational labor. On the other hand, to find $\hat{\beta}(y)$ explicitly, we require that ϵ exist, which leads to the classic problem of "identifiability", and the requirements that $\text{rank}(X) = k$, and $n \geq k$.

For example, in the usual analysis of additive rate relativities, one adds extra constraints on the parameters so that X will be of rank k , and $\hat{\beta}(y)$ will exist [Grimes (1971)]. This is not necessary in a Bayesian regression, so long as n is not so large that Z or z is ill-conditioned. Of course, there may be external reasons, such as economic equity, for using only models in which X has full rank; in this case, one can show that, for "stable" increasing designs, $z \rightarrow I_k$ as $n \rightarrow \infty$ [Jewell (1975)].

The preposterior covariance of the parameter estimation error can be shown to be:

$$(2.9) \quad \phi = V\{\tilde{\beta} - f(\tilde{y})\} = (I_k - ZX)\Delta = (I_k - z)\Delta = (\Delta^{-1} + \epsilon^{-1})^{-1}.$$

Since the precision (inverse covariance) in estimating $\tilde{\beta}$ was Δ^{-1} , prior-to-data, we see that, on the average, a forecast using X increases the precision by ϵ^{-1} ; alternately, ϕ is the " Δ " we expect to have, on the average, as our estimate goes from b to f .

Hachemeister (1974) and Taylor (1974;1975) have both given special versions of (2.6), (2.7). And there are numerous non-Bayesian versions [Theil (1963); Rao (1965)]. However, priority for both forms belongs in the communications theory literature, where generalized least-squares methods have been used for linear (Wiener-Kalman-Bucy) filter estimation problems for many years, (see, for example, Sage and Melsa (1971, pp. 182-4). Further historical remarks are in Jewell (1975).

3. ITERATIVE CALCULATIONS

An interesting feature of credibility regression is the possibility of cascading or serially combining several experiments through recursive calculations. Let X_1, X_2, \dots, X_t be the design matrices for experiments 1, 2, ..., t in which $\tilde{\beta}$ remains the same, but vectors y_1, y_2, \dots, y_t are observed, with known observational error covariance matrices E_{11}, \dots, E_{tt} , for each experiment; the observational dimension n_t may vary from experiment to experiment, but we assume observational independence between experiments, i.e. $C\{\tilde{u}_s; \tilde{u}_t \mid \tilde{\beta}\} = 0$ ($s \neq t$) for all $\tilde{\beta}$.

One possibility for calculation is to combine all experiments into a single large model (1.1), with $y' = [y_1': y_2': \dots]$, $X' = [X_1': X_2': \dots]$, $E = \text{diag}(E_{11}; E_{22}; \dots)$; the dimensions will be $n = \sum n_t$ by k . Even if the second version (2.6) is used, there is a fair amount of simultaneous computation to perform before the single $k \times k$ inversion; furthermore, if the data-gathering is, in fact, sequential in time, then successive forecasts are more and more inefficient.

In Jewell (1975), it is shown that equivalent computations can be performed in the following recursive manner:

- (1) Initialize by defining $b(1) = b$ and $\Delta(1) = \Delta$.
- (2) For period t , assume that current prior moments, $b(t)$ and $\Delta(t)$, are given. Using these and y_t , X_t , and E_{tt} from the current experiment, compute an updated forecast of $\tilde{\beta}$, call it $f_t(y_t)$, from (2.4) or (2.6), and an error covariance ϕ_t from (2.9).
- (3) Continue the computation for period $t + 1$ by setting

$$(3.1) \quad b(t + 1) = f_t(y_t) \quad ; \quad \Delta(t + 1) = \phi_t \quad ;$$

and repeat Step 2.

This iterative process replaces the all-at-once computations by a sequence of smaller ones, with a choice of whether to use an n_t - or k -order inversion at each stage. In fact, if all the E_{tt} are diagonal matrices (each dimension of observation error independent) then one could iterate through every row of all X_t !

It should be emphasized that the final ϕ_t is still a *preposterior* covariance, equal to:

$$(3.2) \quad \phi_t^{-1} = \Delta^{-1} + \sum_{i=1}^t \epsilon^{-1}(i) ; \quad \epsilon^{-1}(i) = X_i' E_{ii}^{-1} X_i ;$$

and is not updated by the y_t . With more specific distributional assumptions, one could in principle update the covariance as well; however, for problems in which control of the variance is appropriate, one would probably use different, nonstationary models, and different techniques, such as Box-Jenkins forecasting, or Wiener-Kalman filtering.

4. SIMPLE TRENDS

Suppose now we believe that the regression parameters are subject to an unknown linear trend, and that, in fact:

$$(4.1) \quad \tilde{\beta} = \tilde{\beta}'_1 + t \cdot \tilde{\beta}'_2, \quad (t = 1, 2, \dots, T)$$

with the model design X held constant. In a direct formulation, one would use $y' = [y'_1 : y'_2 : \dots, y'_T]$, and an $nT \times 2k$ super-design matrix

$$(4.2) \quad \begin{bmatrix} X & X \\ X & 2X \\ \cdot & \cdot \\ \cdot & \cdot \\ \cdot & \cdot \\ X & TX \end{bmatrix} ;$$

there are now $2k$ parameters rearranged in linear format with prior mean vector $b' = [b'_1 : b'_2]$, and prior covariance

$$(4.3) \quad \Delta = \begin{bmatrix} \Delta_{11} & \Delta_{12} \\ \Delta_{21} & \Delta_{22} \end{bmatrix} .$$

Note that it is not reasonable to assume that $\Delta_{12} = \Delta'_{21}$ is void, since most linear models of inflation are proportional in nature, i.e. (4.1) represents

$$(4.4) \quad \tilde{\beta} = (I_k + t\tilde{\Gamma})\tilde{\beta}_1 ,$$

where $\tilde{\Gamma}$ is a scalar or diagonal matrix of unknown inflation rates.

In Section 3, we have shown how computation using (4.2) could be reduced to a series of $n \times 2k$ computations using a design matrix of form $[X:tX]$; for the rest of this section, we further simplify computation using this special structure.

Initialize by defining $b_i(1) = b_i$, and $\Delta_{ij}(1) = \Delta_{ij}$ ($i, j = 1, 2$). Then, if $n = k$, we find two formulae similar to (2.4) for iteration t by using y_t , X_t , E_{tt} , and calculating

$$(4.5) \quad \Delta_1(t) = \Delta_{11}(t) + t\Delta_{12}(t) ; \Delta_2(t) = \Delta_{21}(t) + t\Delta_{22}(t) ; \Delta_0(t) = \Delta_1(t) + t\Delta_2(t) .$$

Inverting one $n \times n$ matrix, we find first

$$(4.6) \quad Z_i(t) = \Delta_i(t)X'(E_{tt} + X\Delta_0(t)X')^{-1} , \quad (i = 1, 2)$$

and then update forecasts of $\tilde{\beta}_1$ and $\tilde{\beta}_2$ through:

$$(4.7) \quad b_1(t+1) = (I_k - Z_1(t)X)b_1(t) - tZ_1(t)Xb_2(t) + Z_1(t)y_t ,$$

$$(4.8) \quad b_2(t+1) = -Z_2(t)Xb_1(t) + (I_k - tZ_2(t)X)b_2(t) + Z_2(t)y_t .$$

On the other hand, if $k < n$, we find $\epsilon^{-1}(t)$ from (3.2), invert one $k \times k$ matrix, and make the following replacements in (4.7), (4.8):

$$(4.9) \quad \left. \begin{array}{l} Z_1(t)X \\ Z_1(t)y_t \end{array} \right\} = \Delta_1(t) \left(I_k + \epsilon^{-1}(t) \Delta_j(t) \right)^{-1} \left\{ \begin{array}{l} \epsilon^{-1}(t) \\ X'E_{tt}^{-1}y_t \end{array} \right. \quad (i = 1, 2)$$

The four components of the preposterior covariance are then updated without further inversion through:

$$(4.10) \quad \Delta_{ij}(t+1) = \Delta_{ij}(t) - Z_1(t)X\Delta_j' \quad (i, j = 1, 2)$$

Thus, in the simple trend case of Bayesian regression, we obtain finally an iterative sequence of calculations, with a single minimal inversion at each step, and separate formulae for updating base values and trends of the unknown model parameters.

REFERENCES

1. Box, G.E.P. and G. C. Tiao, BAYESIAN INFERENCE IN STATISTICAL ANALYSIS, Addison-Wesley, Reading (1973).
2. Grimes, T., "Claim Frequency Analysis," Journal of the Students (U.K. Actuarial) Society, Part III (1971).
3. Hachemeister, "Credibility for Regression Models with Application to Trend," Proceedings, Berkeley Actuarial Research Conference on Credibility, (1974); to appear, Academic Press, New York.
4. Jewell, W. S., "Exact Multidimensional Credibility," Mitteilungen der Vereinigung schweizerischer Versicherungsmathematiker, Vol. 74, No. 2, pp. 193-214 (1974).
5. Jewell, W. S., "Bayesian Regression and Credibility Theory," Research Memorandum, International Institute for Applied Systems Analysis, Laxenburg, Austria (March 1975).
6. Morales, J. A., BAYESIAN FULL INFORMATION STRUCTURAL ANALYSIS, Springer Verlag, Berlin (1971).
7. Rao, C. R., LINEAR STATISTICAL INFERENCE AND ITS APPLICATIONS, Wiley, New York (1965).
8. Sage, A. P. and J. L. Melsa, ESTIMATION THEORY WITH APPLICATIONS TO COMMUNICATIONS AND CONTROL, McGraw-Hill (1971).
9. Taylor, G. C., "Credibility for Time-Heterogeneous Loss Ratios," Proceedings, Berkeley Actuarial Research Conference on Credibility (1974); to appear, Academic Press, New York.
10. Taylor, G. C., "Abstract Credibility," Research Report, Macquarie University, Sydney, and Heriot-Watt University, Edinburgh (February 1975).
11. Theil, H., "On the Use of Incomplete Prior Information in Regression Analysis," J. Amer. Stat. Assoc., Vol. 58, pp. 401-414 (June 1963).
12. Zellner, A., AN INTRODUCTION TO BAYESIAN INFERENCE IN ECONOMETRICS, Wiley, New York (1971).