Journal of Modern Applied Statistical Methods

Volume 10 | Issue 1 Article 24

5-1-2011

Bayesian Regression Analysis with Examples in S-PLUS and R

Sheikh P. Ahmad
University of Kashmir, Srinagar, India, sprvz@yahoo.com

A. A. Khan
University of Kashmir, atharkhan1962@gmail.com

A. Ahmed
Aligarh Muslim University, aqlstat@yahoo.co.in

Journal of Modern Applied Statistical Methods
May 2011, Vol. 10, No. 1, 268-276

1538 – 9472/11/$95.00

Bayesian Regression Analysis with Examples in S-PLUS and R

S. P. Ahmad A. A. Khan
University of Kashmir,
Srinagar, India

A. Ahmed
Aligarh Muslim University,
U.P, India

An extended version of normal theory Bayesian regression models, including extreme-value, logistic and normal regression models is examined. Methods proposed are illustrated numerically; the regression coefficient of *pH* on electrical conductivity (EC) of soil data is analyzed using both *S-PLUS* and *R* software.

Key words: Bayesian regression, extreme-value model, S-PLUS, R.

Introduction

In statistics, regression analysis includes many techniques for modeling and analyzing several variables, when the focus is on the relationship between a dependent variable and one or more independent variables. In practice, many situations involve a heterogeneous population and it is important to consider the relationship of response variable y on concomitant variable x which is explicitly recognized.

One method to examine the relationship of a concomitant variable (or regressor variable) to a response variable y is through a regression model in which y has a distribution that depends upon the regressor variables. This involves specifying a model for the distribution of y given x, where $x = (x_1, x_2, ..., x_p)$ is a $1 \times p$ vector of the regressor variables for an individual.

Let the distribution of y given x be

$$f(y \mid x, \beta, \sigma) = \frac{1}{\sigma} f\left(\frac{y - x\beta}{\sigma}\right), \quad (1.1)$$

where β is a $p \times 1$ vector of regression Coefficients, $\beta = (\beta_1, \beta_2, ..., \beta_p)^T$ and $E[y \mid x, \beta] = x\beta$. The alternative form of (1.1) is

$$y = x\beta + \sigma z \quad (1.2)$$

where

$$z = \frac{y - x\beta}{\sigma}$$

has the standardized distribution with density function *f(z)*. The family of models for which *f(z)* has a standard normal distribution is common in statistical literature (Searle, 1971; Rao, 1973; Seber, 1977; Draper & Smith, 1981; Weisberg, 1985) but models in which z has other distributions belonging to location-scale family (1.2) are also important. For example, extreme value regression models are employed in applications ranging from accelerated life testing (Lawless, 2003; Zelen, 1959) to the analysis of survival data on patients suffering from chronic diseases (Prentice, 1973; Feigl & Zelen, 1965; Krall, et al., 1975).

Furthermore, if data is contaminated with outliers, then the normal distribution can be replaced with Student's *t* distribution (with small degrees of freedom) to have a better fit (e.g., Lange, et al., 1989). Model (1.2) has the ability to accommodate linear as well as non-linear models for the various functional forms of $x\beta$. None of the above authors present a Bayesian approach. Box and Tiao (1973) and Gelman, et al. (1995) discuss this approach of regression

S. P. Ahmad is an Assistant Professor in the Department of Statistics. Email: sprvz@yahoo.com. A. A. Khan is a Professor in the Department of Statistics & O. R. Email: atharkhan1962@gmail.com. A. Ahmed is a Professor in the Department of Statistics. Email: aqlstat@yahoo.co.in.

analysis to deal with normal linear as well as non-linear non-normal models. Zellener (1971) describes Bayesian inference in reference to econometrics, but the discussion is mainly confined to normal linear models. The general framework used for casual inference is presented in Rubin (1974b, 1978a). Bayesian approaches to analyzing regression residuals appear in Zellener (1976), Chaloner and Brant (1988) and Chaloner (1991).

Joint inference for β and *log*σ with Non-informative Prior $p(\beta, \log\sigma)$

Suppose (y_i, x_i) for $i = 1, 2, \ldots, n$ is assumed to be a random sample from location-scale family of models in (1.1) and likelihood is

$$\prod_{i=1}^{n} f(y_i \mid x_i, \beta, \sigma)$$

This implies that

$$l(\beta, \log\sigma) = \sum_{i=1}^{n} \log f(z_i) - n\log\sigma \tag{2.1}$$

where

$$z_i = \frac{y_i - x_i\beta}{\sigma}.$$

Consider the non-informative prior

$$p(\beta, \log\sigma) = 1 \tag{2.2}$$

The joint posterior density of β and *log*σ given data vector $y^T = (y_1, y_2, ..., y_n)$ is

$$p(\beta, \log\sigma \mid x, y) \;\alpha \prod_{i=1}^{n} f(y_i \mid x_i, \beta, \log\sigma)\, p(\beta, \log\sigma) \tag{2.3}$$

where $x = (x_1, x_2, ..., x_n)^T$ is a $n \times p$ matrix of covariates (or regressors) corresponding to response vector *y*. Now joint inference for β and *log*σ can be made from posterior (2.3).

Posterior mode $(\hat{\beta}, \log\hat{\sigma})^T$ of $p(\beta, \log\sigma \mid x, y)$ serves as a point estimate of β and $\log\sigma$. Its calculations require partial derivatives of log posterior

$$\begin{aligned} l^*(\beta, \log\sigma) &= l(\beta, \log\sigma) + \log p(\beta, \log\sigma) \\ &= l(\beta, \log\sigma) \end{aligned} \tag{2.4}$$

Defining partial derivatives as $l^*_\beta = \frac{\partial l^*}{\partial \beta}$, a vector of $(p \times 1)$ partial derivatives,

$l^*_\phi = \frac{\partial l^*}{\partial \phi}$, a scalar and $\phi = \log\sigma$

$l^*_{\beta\phi} = \frac{\partial^2 l^*}{\partial\beta\partial\phi}$, a $(p \times 1)$ vector,

$l^*_{\phi\beta} = \frac{\partial^2 l^*}{\partial\phi\partial\beta}$, a $(1 \times p)$ vector,

$l^*_{\beta\beta} = \frac{\partial^2 l^*}{\partial\beta\partial\beta^T}$, a $(p \times p)$ matrix, and

$l^*_{\varphi\varphi} = \frac{\partial^2 l^*}{\partial\varphi\partial\varphi^T}$.

These derivatives can be defined more explicitly as:

$l^*_\beta = l_\beta$

$l^*_\phi = l_\phi$

$l^*_{\beta\phi} = l_{\beta\phi}$

$l^*_{\phi\beta} = l_{\phi\beta}$

$l^*_{\beta\beta} = l_{\beta\beta}$ and

$l^*_{\phi\phi} = l_{\phi\phi}$.

Consequently, score vector $U(\beta,\phi)$ and Hessian matrix $H(\beta,\phi)$ are a $(p+1)\times 1$ vector

$$U(\beta,\phi)=\begin{bmatrix} l_\beta^* \\ l_\phi^* \end{bmatrix},$$

and a $(p+1)\times(p+1)$ matrix

$$H(\beta,\phi)=\begin{bmatrix} l_{\beta\beta}^* & l_{\beta\phi}^* \\ l_{\phi\beta}^* & l_{\phi\phi}^* \end{bmatrix},$$

therefore, making use of Newton-Raphson iteration scheme, results in posterior mode vector $(\hat{\beta},\hat{\phi})^T$ as

$$\begin{bmatrix} \hat{\beta} \\ \hat{\phi} \end{bmatrix}=\begin{bmatrix} \beta_0 \\ \phi_0 \end{bmatrix}-H^{-1}(\beta_0,\phi_0)\begin{bmatrix} l_\beta^* \\ l_\phi^* \end{bmatrix} \tag{2.5}$$

where $\hat{\phi}=\log\hat{\sigma}$.

The asymptotic posterior covariance matrix of (2.3) can be obtained as

$$\begin{aligned} I^{-1}(\hat{\beta},\hat{\varphi}) &= -H^{-1}(\hat{\beta},\hat{\varphi}) \\ &= \sum(\hat{\beta},\hat{\varphi}). \end{aligned}$$

More clearly, posterior density

$$p(\beta,\phi|x,y)\cong N_{p+1}\left((\hat{\beta},\hat{\phi})^T, I^{-1}(\hat{\beta},\hat{\phi})\right)\left(1+O(n^{-\frac{1}{2}})\right) \tag{2.6}$$

where $N_r(a,b)$ is the r-variate normal distribution with mean vector a and a covariance vector b. This is a first order approximation of the posterior density (e.g., Berger, 1985). An equivalent version of this approximation is the Chi-square approximation, i.e.,

$$W(\beta,\phi)=-2\left[l^*(\beta,\phi)-l^*(\hat{\beta},\hat{\phi})\right]\approx\chi_{p+1}^2.$$

A more accurate approximation, Laplace's approximation (Tierney & Kadane, 1986; Reid, 1988) can be also used, i.e.,

$$p(\beta,\phi\,|\,x,y)\cong (2\pi)^{-\frac{p+1}{2}}\,|\,I(\hat{\beta},\hat{\phi})\,|^{\frac{1}{2}}\,\exp\left[-\frac{1}{2}W(\hat{\beta},\hat{\phi})\right]\left(1+O(n^{-1})\right) \tag{2.7}$$

Any of the approximations can be used both for hypothesis testing and construction of credible regions.

The Marginal Inference for β and ϕ *($\phi = \log\sigma$)*

The marginal densities for β and ϕ are

$$p(\beta\,|\,x,y)=\int p(\beta,\varphi\,|\,x,y)\,d\varphi. \tag{3.1}$$

Similarly, marginal posterior of ϕ can be obtained by

$$p(\varphi\,|\,x,y)=\int p(\beta,\varphi\,|\,x,y)\,d\beta. \tag{3.2}$$

Bayesian analysis is to be based on these two posteriors. For the normal model, $p(\beta\,|\,x,y)$ and $p(\phi\,|\,x,y)$ can be obtained in closed form (e.g., Zellener, 1971). However, for non-normal members of location-scale family, these marginals can be obtained through numerical integration only (e.g. Naylor & Smith, 1982). The alternative approach is to deal with asymptotic theory approach (e.g., Tierney, Kass & Kadane, 1989a; Leonard, et al., 1989). Normal and Laplace's approximations can be written directly for posterior densities $p(\beta\,|\,x,y)$ and $p(\phi\,|\,x,y)$ as under:

a) Normal Approximation:

Marginal posterior density of β can be approximated by normal distribution, i.e.,

$$p(\beta \mid x, y) \cong N_p(\hat{\beta}, I_{11}^{-1}) \quad (3.3)$$

where $\hat{\beta}$ is the posterior mode and I_{11}^{-1} is a $p \times p$ matrix defined as

$$I^{-1}(\hat{\beta}, \hat{\phi}) = \begin{bmatrix} I_{11}^{-1} & I_{12}^{-1} \\ I_{21}^{-1} & I_{22}^{-1} \end{bmatrix}$$

where $\hat{\phi} = \log \hat{\sigma}$ and suffixes 1 and 2 to I stand for $\hat{\beta}$ and $\hat{\phi}$, respectively. This approximation is equivalent to the Chi square approximation defined as:

$$(\beta - \hat{\beta})^T I_{11}(\beta - \hat{\beta}) \approx \chi_p^2 .$$

Corresponding approximations for $p(\phi \mid x, y)$ can be written as:

$$p(\phi \mid x, y) = N_1(\hat{\phi}, I_{22}^{-1}) \qquad (3.4)$$

This is equivalent to the Chi square approximation, i.e.,

$$(\varphi - \hat{\varphi})^T I_{22}(\varphi - \hat{\varphi}) \approx \chi_1^2 .$$

b) Laplace's Approximation:

Laplace's approximation can also be used to approximate marginal density of β, i.e.,

$$p(\beta \mid y) \cong \left[\frac{| I(I(\hat{\beta}, \hat{\phi})) |}{2\pi \, | I(\beta, \hat{\phi}(\beta)) |} \right]^{\frac{1}{2}} \exp[l^*(\beta, \hat{\phi}(\beta)) - l^*(\hat{\beta}, \hat{\phi})] \quad (3.5)$$

where $\hat{\phi}(\beta)$ is the posterior mode of ϕ for a fixed β.

Corresponding approximation for $p(\phi \mid x, y)$ can also be written as

$$p(\phi \mid x, y) \cong (2\pi)^{-\frac{p}{2}} \left[\frac{| I(\hat{\beta}, \hat{\phi}) |}{| I(\beta, \hat{\phi}(\beta)) |} \right]^{\frac{1}{2}} \exp[l^*(\hat{\beta}(\phi), \phi) - l^*(\hat{\beta}, \hat{\phi})] \quad (3.6)$$

where $\hat{\beta}(\phi)$ is the posterior mode of β for a fixed ϕ.

Bayesian Regression Analysis of the Extreme-Value Model

Let y be the response vector and x_i be the vector for the i[th] observation. Assume that

$$z_i = \frac{y_i - x_i^T \beta}{\sigma} \sim f \qquad (4.1)$$

for some f (extreme value distribution). Consequently, in terms of general notation $\theta = (\beta, \sigma)^T$, a vector of length $(\beta + 1)$ and likelihood is given by:

$$\prod_{i=1}^{n} f(y_i \mid x_i, \beta, \sigma) .$$

This implies that

$$l(\beta, \sigma) = \log \prod_{i=1}^{n} f(y_i \mid x_i, \beta, \sigma) = \sum_{i=1}^{n} \log f(z_i) - n \log \sigma \qquad (4.2)$$

where z_i is defined in (4.1).

Taking partial derivatives with respect to μ and σ results in

$$l_\beta = \frac{\partial l}{\partial \beta} = -\frac{1}{\sigma} \sum_{i=1}^{n} (1 - e^{z_i}) x_i^T$$

$$l_\sigma = \frac{\partial l}{\partial \sigma} = -\frac{1}{\sigma}\sum_{i=1}^{n} z_i(1-e^{z_i}) - \frac{n}{\sigma}$$

$$l_{\beta\sigma} = \frac{\partial^2 l}{\partial\beta\partial\sigma} = -\frac{1}{\sigma^2}\sum_{i=1}^{n}(z_i e^{z_i} + e^{z_i} - 1)x_i^T$$

$$l_{\sigma\beta} = \frac{\partial^2 l}{\partial\sigma\partial\beta} = -\frac{1}{\sigma^2}\sum_{i=1}^{n}(z_i e^{z_i} + e^{z_i} - 1)x_i^T$$

$$l_{\beta\beta} = \frac{\partial^2 l}{\partial\beta\partial\beta^T} = -\frac{1}{\sigma^2}\sum_{i=1}^{n} e^{z_i} x_i^T x_i$$

and

$$l_{\sigma\sigma} = \frac{\partial^2 l}{\partial\sigma\partial\sigma^T} = -\frac{1}{\sigma^2}\sum_{i=1}^{n}(z_i^2 e^{z_i} + 2z_i e^{z_i} - 2z_i) + \frac{n}{\sigma^2}.$$

Following the standard approach of Box and Tiao (1973) and Gelman, et al. (1995), assuming the prior

$$p(\beta,\sigma) \cong p(\beta)p(\sigma) \tag{4.3}$$

where $p(\beta)$ and $p(\sigma)$ are priors for β and σ. Using Bayes theorem obtain the posterior density $p(\beta,\sigma \mid y)$ is obtained as

$$p(\beta,\sigma \mid x,y) \;\alpha\; \prod_{i=1}^{n} f(y_i \mid x_i,\beta,\sigma)\, p(\beta,\sigma) \tag{4.4}$$

The log-posterior is given by

$$\log p(\beta,\sigma \mid x,y) = \log\prod_{i=1}^{n} p(y_i \mid x_i,\beta,\sigma) + \log p(\beta) + \log p(\sigma)$$

or

$$l^*(\beta,\sigma) = l(\beta,\sigma) + \log p(\beta) + \log p(\sigma). \tag{4.5}$$

For a prior $p(\beta,\sigma) \cong p(\beta)p(\sigma) = 1$, $l^*_\beta = l_\beta$, $l^*_\sigma = l_\sigma$, $l^*_{\beta\sigma} = l_{\beta\sigma}$, $l^*_{\sigma\beta} = l_{\sigma\beta}$, $l^*_{\beta\beta} = l_{\beta\beta}$ and $l^*_{\sigma\sigma} = l_{\sigma\sigma}$. The posterior mode is obtained by maximizing (4.5) with respect to β and σ. The score vector of the log posterior is given by

$$U(\beta,\sigma) = (l^*_\beta, l^*_\sigma)^T$$

and the Hessian matrix of log posterior is

$$H(\beta,\sigma) = \begin{bmatrix} l^*_{\beta\beta} & l^*_{\beta\sigma} \\ l^*_{\sigma\beta} & l^*_{\sigma\sigma} \end{bmatrix}.$$

Posterior mode $(\hat{\beta},\hat{\sigma})$ can be obtained from iteration scheme

$$\begin{bmatrix} \hat{\beta} \\ \hat{\sigma} \end{bmatrix} = \begin{bmatrix} \beta_0 \\ \sigma_0 \end{bmatrix} - H^{-1}(\beta_0,\sigma_0)\begin{bmatrix} l^*_\beta \\ l^*_\sigma \end{bmatrix}. \tag{4.6}$$

Consequently, the modal variance Σ can be obtained as

$$I^{-1}(\hat{\beta},\hat{\sigma}) = -H^{-1}(\hat{\beta},\hat{\sigma}).$$

Using the normal approximation, a bivariate normal approximation of $p(\beta,\sigma \mid x,y)$ can be directly written as:

$$p(\beta,\sigma \mid y) \cong N_{p+1}\left((\hat{\beta},\hat{\sigma})^T, I^{-1}(\hat{\beta},\hat{\sigma})\right)\left(1 + O(n^{-\frac{1}{2}})\right). \tag{4.7}$$

Similarly, a Bayesian analog of likelihood ratio criterion can be written as:

$$W(\beta,\sigma) = -2[l^*(\mu,\sigma) - l^*(\hat{\mu},\hat{\sigma})] \approx \chi^2_{p+1} \quad (4.8)$$

Using Laplace's approximation, $p(\beta,\sigma \mid x,y)$ can be written as:

$$p(\beta,\sigma \mid x,y) \cong (2\pi)^{-\frac{p-1}{2}} \mid I(\hat{\beta},\hat{\sigma}) \mid^{\frac{1}{2}} \exp[l^*(\mu,\sigma) - l^*(\hat{\mu},\hat{\sigma})]\left(1+O(n^{-1})\right). \quad (4.9)$$

The marginal Bayesian inference about β and σ is based on the marginal posterior densities of these parameters. Marginal posterior for β can be obtained after integrating out $p(\beta,\sigma \mid x,y)$ with respect to σ,

$$p(\beta \mid x,y) = \int p(\beta,\sigma \mid x,y)\, d\sigma. \quad (4.10)$$

Similarly, the marginal posterior of σ can be obtained by:

$$p(\sigma \mid x,y) = \int p(\beta,\sigma \mid x,y) d\beta. \quad (4.11)$$

The normal approximation for marginal posterior $p(\beta \mid x,y)$ can be written as:

$$p(\beta \mid x,y) = N_p(\hat{\beta}, I_{11}^{-1}) \quad (4.12)$$

where $\hat{\beta}$ is the posterior mode and I_{11}^{-1} is a $(p \times p)$ matrix defined as

$$I^{-1}(\hat{\beta},\hat{\sigma}) = \begin{bmatrix} I_{11}^{-1} & I_{12}^{-1} \\ I_{21}^{-1} & I_{22}^{-1} \end{bmatrix}.$$

The Bayesian analog of likelihood ratio criterion can also be defined as a test criterion as:

$$(\beta - \hat{\beta})^T I_{11}(\beta - \hat{\beta}) \approx \chi^2_p. \quad (4.13)$$

Laplace's approximation of marginal posterior density $p(\beta \mid x,y)$ can be given by:

$$p(\beta \mid x,y) \cong \left[\frac{\mid I(\hat{\beta},\hat{\sigma}) \mid}{2\pi \mid I(\beta,\hat{\sigma}(\beta)) \mid}\right]^{\frac{1}{2}} \exp[l^*(\beta,\hat{\sigma}(\beta)) - l^*(\hat{\beta},\hat{\sigma})]. \quad (4.14)$$

Similarly, $p(\sigma \mid x,y)$ can be approximated and results corresponding to normal and Laplace's approximation can be written as

$$p(\sigma \mid x,y) = N_1(\hat{\sigma}, I_{22}^{-1}) \quad (4.15)$$

or equivalently,

$$(\sigma - \hat{\sigma})^T I_{22}(\sigma - \hat{\sigma}) \approx \chi^2_1 \quad (4.16)$$

$$p(\sigma \mid x,y) \cong (2\pi)^{-\frac{p}{2}} \left[\frac{\mid I(\hat{\beta},\hat{\sigma}) \mid}{\mid I(\hat{\beta}(\sigma),\sigma) \mid}\right]^{\frac{1}{2}} \exp[l^*(\hat{\beta}(\sigma),\sigma) - l^*(\hat{\beta},\hat{\sigma})]. \quad (4.17)$$

Numerical Illustrations

Numerical illustrations are implemented in S-PLUS software for Bayesian regression analysis. These illustrations are show the strength of Bayesian methods in practical situations. Soil samples were collected from rice growing areas as well as fruit orchids of Kashmir valley and were analyzed for some relevant parameters. In our work, we studied pH and E.C in the soil of Kashmir valley. The functions *survReg* and *cendnsorReg* were used for Bayesian analysis of various regression models with non-informative prior. S-PLUS has a function *censorReg* for regression analysis; this has a very substantial overlap with *survReg* but is more general in that it allows truncation as well as censoring (Venables & Ripley, 2002). The usage of *survReg* and *censorReg* are:

survReg(formula, data, dist)

censorReg(formula, data, dist)

where

- *formula:* a formula expression as for other regression models;
- *data:* optional data frame in which to interpret the variable occurring in the formula; and
- *dist:* assumed distribution for y variable.

References

Berger, J. O. (1985). *Statistical decision theory and bayesian analysis, 2nd Edition*. New York: Springer-Verlag.

Box, G. E. P., & Tiao, G. C. (1973). *r*

Chaloner, K. (1991). Bayesian residual analysis in the presence of censoring. *Biometrika*, *78*, 637-644.

Chaloner, K., & Brant, R. (1988). A Bayesian approach to outlier detection and residual analysis. *Biometrika*, *75*, 651-659.

Draper, N. R., & Smith, H., Jr. (1981). *Applied regression analysis*. New York: Wiley.

Feigl, P., & Zelen, M. (1965). Estimation of exponential survival probabilities with concomitant information. *Biometrics*, *21*, 826-836.

Gelman, A., Carlin, J. B., Stern, H. S., & Rubin, D. B. (1995). *Bayesian data analysis*. London: Chapman and Hall.

Krall, J., Uthoff, V., & Harley, J. (1975). A step-up procedure for selecting variables associated with survival. *Biometrics*, *31*, 49-57.

Lange, K. L., Little, R. J. A., & Taylor, J. M. G. (1989). Robust statistical modeling using the t distribution. *Journal of the American Statistical Association*, *84*, 881-896.

Lawless, J. F. (2003). *Statistical models and methods for lifetime data*. New York: Wiley.

Leonard, T., Hsu, J. S. J., & Tsui, K. W. (1989). Bayesian marginal inference. *Journal of the American Statistical Association*, *84*, 1051-1058.

Naylor, J. C., & Smith, A. F. M. (1982). Applications of a method for efficient computation of posterior distributions. *Applied Statistics*, *31*, 214-225.

Prentice, R. L. (1973). Exponential survival with censoring and explanatory variables. *Biometrika*, *60*, 279-288.

Rao, C. R. (1973). *Linear statistical inference and its applications*. New York: Wiley.

Reid, N. (1988). Saddle point methods and statistical inference. *Statistical Science*, *3*, 213-238.

Rubin, D. B. (1974b). Estimating casual effects of treatments in randomized and nonrandomized studies. *Journal of Educational Psychology*, *66*, 688-701.

Rubin, D. B. (1978a). Bayesian inference for casual effects: the role of randomization. *The Annals of Statistics*, *6*, 34-58.

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

Seber, G.A.F. (1977). *Linear regression analysis*. New York: Wiley.

Tierney, L., & Kadane, J. B. (1986). Accurate approximations for posterior moments and marginal densities. *Journal of the American Statistical Association*, *81*, 82-86.

Tierney, L., Kass, R. E., & Kadane, J. B. (1989a). Fully exponential Laplace approximations to the expectations and variances of nonpositive functions. *Journal of the American Statistical Association*, *84*, 710-716.

Venables, W. N., & Ripley, B. D. (2002). *S programming*. New York: Springer-Verlag.

Weisberg, S. (1985). *Applied linear regression*. New York: Wiley.

Zelen, M. (1959). Factorial experiments in life testing. *Technometrics*, *1*, 269-288.

Zellener, A. (1971). *An introduction to bayesian inference in econometrics*. New York: Wiley.

Zellener, A. (1976). Bayesian and non-Bayesian analysis of the regression model with multivariate Student-t error terms. *Journal of the American Statistical Association*, *71*, 400-405.

Table 1: A Summary of Derivatives of Log-Likelihoods

Derivatives	Distributions		
	Normal	Extreme-Value	Logistic
l_β	$\frac{1}{\sigma}\sum_{i=1}^{n} z_i x_i^T$	$-\frac{1}{\sigma}\sum_{i=1}^{n}(1-e^{z_i})x_i^T$	$\frac{1}{\sigma}\sum_{i=1}^{n}\left(\frac{e^{z_i}-1}{e^{z_i}+1}\right)x_i^T$
l_σ	$\frac{1}{\sigma}\sum_{i=1}^{n} z_i^2 - \frac{n}{\sigma}$	$-\frac{1}{\sigma}\sum_{i=1}^{n} z_i(1-e^{z_i}) - \frac{n}{\sigma}$	$\frac{1}{\sigma}\sum_{i=1}^{n} z_i\left(\frac{e^{z_i}-1}{e^{z_i}+1}\right) - \frac{n}{\sigma}$
$l_{\beta\sigma}$	$-\frac{2}{\sigma^2}\sum_{i=1}^{n} z_i x_i^T$	$-\frac{1}{\sigma^2}\sum_{i=1}^{n}(z_i e^{z_i} + e^{z_i} - 1)x_i^T$	$-\frac{1}{\sigma^2}\sum_{i=1}^{n}\left(\frac{e^{2z_i} + 2z_i e^{z_i} - 1}{(e^{z_i}+1)^2}\right)x_i^T$
$l_{\sigma\beta}$	$-\frac{2}{\sigma^2}\sum_{i=1}^{n} z_i x_i^T$	$-\frac{1}{\sigma^2}\sum_{i=1}^{n}(z_i e^{z_i} + e^{z_i} - 1)x_i$	$-\frac{1}{\sigma^2}\sum_{i=1}^{n}\left(\frac{e^{2z_i} + 2z_i e^{z_i} - 1}{(e^{z_i}+1)^2}\right)x_i$
$l_{\beta\beta}$	$-\frac{2}{\sigma^2}\sum x_i^T x_i$	$-\frac{1}{\sigma^2}\sum_{i=1}^{n} e^{z_i} x_i^T x_i$	$-\frac{2}{\sigma^2}\sum_{i=1}^{n}\left(\frac{e^{z_i} x_i^T x_i}{(e^{z_i}+1)^2}\right)$
$l_{\sigma\sigma}$	$-\frac{3}{\sigma^2}\sum_{i=1}^{n} z_i^2 + \frac{n}{\sigma^2}$	$-\frac{1}{\sigma^2}\sum_{i=1}^{n}(z_i^2 e^{z_i} + 2z_i e^{z_i} - 2z_i) + \frac{n}{\sigma^2}$	$-\frac{2}{\sigma^2}\sum_{i=1}^{n} z_i\left(\frac{e^{2z_i} + z_i e^{z_i} - 1}{(e^{z_i}+1)^2}\right) + \frac{n}{\sigma^2}$

where $z_i = \frac{y_i - x_i\beta}{\sigma}$, $\quad i = 1,2,\ldots,n$.

Table 2: A Summary of Prior Densities for Location Parameter β

Name of Density	$p(\mu)$	$\frac{p'(\mu)}{p(\mu)}$	$\left[\frac{p'(\mu)}{p(\mu)}\right]'$
Non-Informative	Constant	Zero	Zero
Normal	$c\exp\left(-\frac{1}{2}D^T D\right)$	$-\frac{D}{\sigma_0}$	$-\frac{1}{\sigma_0^2}I$

Where $D^T = (D_1, D_2, \ldots, D_p)$ a $(p \times 1)$ vector, $D_i = \frac{\beta_i - \beta_{0i}}{\sigma_{oi}}$ for $i = 1,2,\ldots,p$; I stands for identity $(p \times p)$ matrix and c is the normalizing constant.

Table 3: Regression Coefficient of pH on EC for Various Models

Regression Model	$\hat{\beta}$ (Intercept)		Posterior Std.	
	$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\beta}_0$	$\hat{\beta}_1$
Extreme-Value	6.71	2.32	0.0447	0.2881
Logistic	6.29	2.41	0.0365	0.2332
Normal	6.33	2.00	0.0335	0.1890

Table 4: Approximate Normal Posterior Quantiles for Regression Coefficient of Various Models

Model	Posterior	Posterior Quantile					
		0.025	0.25	0.50	0.75	0.95	0.975
Extreme-Value	Normal	6.63	6.68	6.72	6.75	6.79	6.80
Logistic	Normal	6.16	6.20	6.23	6.25	6.29	6.30
Normal	Normal	6.26	6.31	6.33	6.35	6.39	6.40