CHAPTER 4

REGRESSION ANALYSIS FOR CENSORED DATA

Introduction

Regression analysis for censored data has been a main focus for many research activities. The models considered by different authors are experimental models in which the independent variables $X_i$ are under control and are equally spaced. At each level of $X_i$, an equal number $k$ of observations of independent variable $Y_{ij}$ are made. The general regression model is

$$Y_{ij} = \alpha + \beta_i (x_i - \bar{x}) + e_{ij} \quad \ldots \ldots \quad (4.1.1)$$

$$i = 1, 2, \ldots, m ; \quad j = 1, 2, \ldots, k$$

where $\alpha$ and $\beta$ are intercept and slopes respectively. The error quantities $e_{ij}$ are assumed to be independent and have identical normal distribution with $N(0, \sigma^2)$.

The estimation procedures applied by different authors to find the unknown parameters based on censored regression sample are:

(i) Maximum likelihood method

(ii) Partial likelihood method

(iii) Sequential procedures

(iv) Least Squares method and

(v) Iterative techniques.

One can apply maximum likelihood techniques to obtain regression estimators, which are known to be in general consistent, efficient and asymptotically normal. There are many examples in literature of inconsistency of ML estimators. Neymann and Scott [1948], Basu [1955], Kiefer and Wolfowitz [1956], Bahadur [1958] and Hannan [1960] gave few examples where ML method leads to inconsistent estimators. The important feature of consistent cases is that by increasing the sample size one can separate or
distinguish between distributions having different parameter points by considering the likelihood function. When the sample is full and when the condition of identically distributed independent random variables are emphasized, the ML method gives satisfactory results. But in case of censored sample observations, the likelihood function may become complicated and the associated equations may sometimes pose multiple roots.

To avoid these difficulties, Cox [1972] applied partial likelihood methods to estimate the parameters of $\alpha$ and $\beta_1$ of the model (4.1.1) and found that these estimators are flexible and robust. Breslow [1969] Jones and Whitehead [1979], Tsiatis [1982], Slud[1984], Gu and Lai [1991] and Gu and Ying[1993] proposed various types of sequential procedures to estimate the unknown parameters of regression models based on full sample observations. When there is no censoring, this approach gives equally good results when compared to classical least squares method. Attempts have been made to extend this approach to censored data by Miller [1976], Buckley and James [1979], Koul, Susarla and Van Ryzin [1981], Miller and Halperin [1982] and Lai and Ying [1991]. Miller's estimators do not possess the property of consistency, where as Koul et al's method requires the condition that censoring variables to be identically distributed independent (iid) random variables.

The least squares estimation procedure of finding the unknown parameters of regression model [4.1.1] was first studied by Chen and Dixon[1972]. They considered two contaminated models namely, models with location and scalar errors. Contamination error is assumed to occur with certain probability for each time an observation is made. A location error occurs when the error variable has a probability $P > 0$ of coming from $N(\lambda \sigma, \sigma^2)$ with $-\infty < \lambda < \infty$ and $\lambda \neq 0$. A scalar error occurs when the error variable has probability $P > 0$ of coming from $N(0, \lambda^2 \sigma^2)$ with $\lambda > 0$ and $\lambda \neq 1$. They observed that winsorization and trimming cause some loss in efficiency when the samples come from the uncontaminated model described in (4.1.1). They also found that the least squares estimators of $\alpha$ and $\beta_1$ are less variable in winsorization than in trimming. In contaminated models, especially the models with a scalar error, a sample from linear regression
analysis will benefit if it is either winsorized or trimmed in order to reduce the mean square errors of both $\alpha$ and $\beta$.

Leurgans (1987) obtained a different extension of least squares estimators, but justification of her approach requires that the censoring observations are to be i.i.d random variables and are independent of covariates.

For regression models with incomplete data, the least squares techniques cannot be directly applied without first correcting the bias inherent in the missing data. The pattern of missing data is usually assumed to be known. Though the estimation obtained by least squares techniques are unbiased, Breiman, Tsur and Zemel (1992) proposed an iterative algorithm to improve the efficiencies of these estimators. This algorithm is based on a simple idea: First, fill the missing data using predictors based on the available observations and using this data, the improved estimators are obtained by applying least squares method as if no data is missing. These two steps are iterated until the procedure converges. This algorithm results as good estimators as maximum likelihood estimators when errors are normally distributed. Similar procedures were already proposed by Schmee and Hahn (1979), Buckley and James (1979), James and Smith (1984), Chatterjee and McLeish (1986), Ritov (1990) and Tsur and Zemel (1990).

This chapter deals with the Least Squares procedure of estimating the regression coefficients, when the sample observations are proportionally censored. The optimum choices of $\alpha$ and $\beta$, variances and covariances of the estimators of $\alpha$ and $\beta$ are explained. The generalized least squares techniques are explained in case of stochastic regression models. The effect of autocorrelation in the data, some conventional tests of autocorrelation coefficient, the improved estimators of regression coefficients by using iterative techniques and the confidence intervals for these estimators are also dealt with.
4.2 Regression Estimators

The regression model under study is

$$Y_{ij} = \alpha + \beta_i x_{ij} + e_{ij} \quad \ldots \ldots \quad (4.2.1)$$

where $i = 1, 2, \ldots, m$; $j = 1, 2, \ldots, k$.

Here $x_j$ are independent variables and $Y_{ij}$ are dependent variables and the error term $e_{ij}$ is assumed to be normally distributed with mean zero and variance $\sigma^2$. A random sample of size $n (=mk)$ from the above model is denoted as $(X_i, Y_{ij})$. The $j$th order statistics of $Y_{i1}, Y_{i2}, \ldots, Y_{ik}$ be $Y_{i(j)}$ where

$$Y_{i(1)} \leq Y_{i(2)} \leq \ldots \leq Y_{i(k)}.$$ If the censoring is symmetric (at $r$th level), then

$$Y_{i(w)} = \frac{(\lfloor kp \rfloor + 1) Y_{i([kp]+1)} + \sum_{j=([kp]+1)}^{(k-[kp])} Y_{i(j)} + (k-[kp]) Y_{i(k-[kp])}}{k} \quad \ldots \quad (4.2.2)$$

and

$$Y_{i(t)} = \frac{\sum_{j=(kp+1)}^{(k-[kp])} Y_{i(j)}}{K-2[kp]} \quad \ldots \quad (4.2.3)$$

where $r = \lfloor kp \rfloor$.

Thus at $r$th level winsorization, the sample includes the observations

$$\{X_i, Y_{i(1)} = Y_{i(2)} = \ldots = Y_{i(r+1)} = Y_{i(r+2)} = \ldots = Y_{i(k-1)} = \ldots = Y_{i(k-r)}\}$$

denoted as $(X_i, Y_i)$

Similarly at $r$th level trimming, the sample includes the observations

$$\{X_i, Y_{i(r+1)}), \ldots, (X_i, Y_{i(k-r)})\}$$

Here we can determine the coefficients $\beta_i$ by minimizing the error sum of squares.
Let us write

$$\bar{Y}_i = Y_{i(t)} + \delta (Y_i - Y_{i(t)})$$

where $E[\delta (Y_i - Y_{i(t)})] = 0$.

The expression (4.2.4) can be written as

$$b^2 + \beta \cdot c \beta - 2 \beta \cdot \sigma_0 + \text{var}(y).$$

Here $b = \alpha - E(Y_i) + \beta_1 E(x_1) + \ldots + \beta_m E(x_m)$

The optimum choice of $\beta$ and $b$ are

$$b = 0 \quad \text{and} \quad c\beta = \sigma_0 \quad \text{(or)} \quad \hat{\beta} = c^{-1} \sigma_0$$

where $\sigma_0$ is the vector variances of $\{Y_i\}$ with respect to $x_1, x_2, \ldots$ and $C$ is the variance-covariance Matrix of $x_1, x_2, \ldots$

$$\therefore \quad \beta_i^\wedge = \frac{\sum (x_i - \bar{x})(y_i - \bar{y}_i)}{\sum (x_i - \bar{x})^2}$$

$$\text{Var} (\beta_i) = c^{-1} \sigma^2$$

$$= \frac{\sigma^2}{\sum (x_i - \bar{x})^2}$$

Here $\sigma^2$ is the dispersion matrix of $e_{ij}$ and the optimum value of $\alpha$ is

$$\hat{\alpha} = \text{Min} [\alpha_i : \alpha_i = \bar{y}_i - \hat{\beta}_i E(X_i)]$$

$$\text{with} \quad \text{var} (\hat{\alpha}) = \frac{\sigma^2 \sum x_i^2}{\sum (x_i - \bar{x})^2}$$
\[ \text{Cov}(\hat{\alpha}, \hat{\beta}) = \frac{-\sigma^2 \sum x_i}{h \sum (x_i - \bar{x})^2} \]

where \( h = k - 2[kp] \)

To minimize \( \text{var}(\hat{\beta}) \), one can choose \( x_i \) such that \( \sum (x_i - \bar{x})^2 \) is as large as possible. To make \( \text{var}(\hat{\alpha}) \) minimum, one can choose \( x_i \) such that

\[ \frac{\sum x_i^2}{\sum (x_i - \bar{x})^2} \]

is as small as possible.

Since \( \sum (x_i - \bar{x})^2 \leq \sum x_i^2 \) the variance of \( \hat{\alpha} \) is minimum if \( x_i \) are chosen from origin. This leads \( \text{Cov}(\hat{\alpha}, \hat{\beta}) = 0 \)

To estimate \( \sigma^2 \), we can generally use one of the following formulae

\[ \hat{\sigma}^2 = \frac{1}{h-2} (Y'Y - Y'X c^{-1}c_0) \]

\[ = \frac{1}{h-2} (Y'Y - \hat{\beta}'X'Y) \]

4.3. Stochastic Regression Estimators

Consider a stochastic regression model as

\[ Y_{i(j)} = \alpha + \beta_1 x_i + e_{i(j)} \quad \ldots \quad (4.3.1) \]

\[ i = 1, 2, \ldots, m; \quad j = 1, 2, \ldots, k. \]

where \[ e_{i(j)} = p e_{i(j-1)} + e_{i(j)} \quad \ldots \quad (4.3.2) \]

and \[ E[e_{i(j)} e_{i(j+s)}] = \sigma_{ii}^2 \quad \text{for} \quad s = 0 \]

\[ = 0 \quad \text{for} \quad s \neq 0. \]

Here \( p \) is called first order auto correlation coefficient.

From (4.3.2)
This gives
\[ \sigma^2_{ii} = \frac{\sigma^2(\varepsilon)}{1 - \rho^2}. \]

Further
\[ E(e_{i(j-1)}^2) = \rho \sigma^2_{ii}(\varepsilon) \]
\[ E(e_{i(j-2)}^2) = \rho^2 \sigma^2_{ii}(\varepsilon) \]
and in general
\[ E(e_{i(j)} e_{i(j-l)}) = \rho^l \sigma^2_{ii}(\varepsilon). \]

Assuming that the censoring is symmetric at rth level, from the model (4.2.1), the least squares estimator for \( \beta \) of general linear regression model is
\[ \hat{\beta} = (x'x)^{-1} x'y \]
\[ = \beta + (x'x)^{-1} x'\varepsilon \]
which gives that \( E(\hat{\beta}) = \beta \); and var \( (\hat{\beta}) = \sigma^2 (x'x)^{-1} \).

The variance-covariance matrix of \( e_{i(j)} \) at rth level censoring for the model (4.3.1) is
\[
\text{E} (e_i e_j) = \sigma^2_{ii} \begin{pmatrix}
1 & \rho & \rho^2 & \ldots & \rho^{h-1} \\
\rho & 1 & \rho & \ldots & \rho^{h-2} \\
\vdots & \vdots & \vdots & \ddots & \vdots \\
\rho^{h-1} & \rho^{h-2} & \rho^{h-3} & \ldots & 1
\end{pmatrix} = \sum \ldots \quad (4.3.6)
\]

which gives, by generalized least squares techniques,

\[
\text{Var}(\hat{\beta}) = E \left[ (\hat{\beta} - \beta) (\hat{\beta} - \beta)^\prime \right]
\]

\[
= (x'x)^{-1} x' \sum x (x'x) \quad \ldots \ldots \quad (4.3.7)
\]

\[
= \frac{1}{k-[kp]} \sum_{i=[(kp)+1]}^{(k-[kp])} \left[ x_i \right] [x_i]^\prime \quad \ldots \ldots \quad (4.3.8)
\]

\[
\begin{align*}
\sigma^2 = \frac{1}{(k-[kp])} \sum_{i=[(kp)+1]}^{(k-[kp])} (x_i - \bar{x})^2 \\
M = \frac{1}{\sum (x_i - \bar{x})^2} \left( 1 + \frac{1}{2p} \frac{Cov(x_i, x_{i+1})}{\text{Var}(x_i)} + \frac{2}{\text{Var}(x_i)} + \ldots + \frac{2^{h-1}}{\text{Var}(x_i)} \right)
\end{align*}
\]

where \( M \) is a non-negative term.

In the presence of auto correlation between error terms, the ordinary least squares principle under estimates the regression coefficients.
If we define \( x_{ij} = \rho x_{i(j-1)} + \varepsilon_{ij} \)

then \( 1 + M = (1 + 2\rho^2 + 2\rho^4 + \ldots + 2\rho^{2(h-1)}) \)

\[
= \frac{1 + \rho^2}{1 - \rho^2} \quad \text{for large } h.
\]

Thus when the potentiality of sample size is high or in case of limited censoring

\[
\text{Var}(\hat{\beta}) = \frac{\sigma^2}{\sum (x_i - \bar{x})^2 \left( \frac{1 + \rho^2}{1 - \rho^2} \right)} \quad \ldots \quad (4.3.9)
\]

4.4. Some Conventional Tests for Autocorrelation Coefficient:

Auto correlated disturbances pose a serious problem for the use of least squares techniques. For large samples, different tests may be applied to the computed residuals from least squares regression. These residuals are

\[
e = y - x \hat{\beta} = \varepsilon + x (\beta - \hat{\beta})
\]

The most useful theoretical test statistic, known as the Von Neumann ratio statistic, is defined as

\[
\frac{\delta^2}{s^2} = \frac{\sum_{j=2}^{h} (e_j - e_{j-1})^2 / h - 1}{\sum_{j=1}^{h} (e_j - \bar{e})^2 / h} \quad \ldots \quad (4.4.1)
\]

This is the ratio of mean square successive difference to the variance. For large sample size, \( \delta^2/s^2 \) may be taken as approximately normal with

\[
\mathbb{E} (\delta^2/s^2) = \frac{2h}{h - 1} \quad \text{and}
\]

\[
\text{Var} (\delta^2/s^2) = \frac{4h^2(h-2)}{(h+1)(h-1)^3}
\]
When the sample size is small, Durbin and Watson investigated the sampling distribution of the statistic, known as Durbin-Watson 'd' statistic as

\[
d = \frac{\sum_{j=2}^{h} (e_j - e_{j-1})^2}{\sum_{j=1}^{h} e_j^2}
\]  

(4.4.2)

This statistic is related to the Von Neumann ratio by

\[
d = \frac{\delta^2}{s^2} \left( \frac{h-1}{h} \right).
\]

Since the sampling distribution of \(d\) is depending upon the values of \(x\), Durbin and Watson established the upper (\(d_u\)) and lower (\(d_L\)) limits for \(d\) at various significance levels. For testing the hypothesis \(H_0: \rho = 0\),

- reject if \(d < d_L\); do not reject if \(d > d_u\).
- If \(d_L < d < d_u\), the test is considered as inconclusive.

**Theorem:** Let \(W\) be \(h \times h\), symmetric non-negative positive definite matrix: Then the unbiased estimate \(\hat{\beta}\), which is linear in \(Y\) and minimizes the quadratic form

\[
E (\hat{\beta} - \beta)' W (\hat{\beta} - \beta) \text{ is again given by} 
\]

(4.4.3)

\[
\hat{\beta} = (X' \sum^{-1} X)^{-1} X' \sum^{-1} Y. 
\]  

(4.4.4)

**Proof:** Since \(W\) is symmetric, positive definite Matrix, it has multiple eigen values. Then (4.4.3) can be written as

\[
\sum_{j=1}^{h} \lambda_j q_j' E (\hat{\beta} - \beta) (\hat{\beta} - \beta)' q_j 
\]

(4.4.5)
Where \( q_j \) is the right eigen vector corresponding to eigen value \( \lambda_j \). By definition of \( W \), the eigen values \( \lambda_j \) are all real and non-negative. Since the estimator \( \hat{\beta} \) is linear in \( Y \), write \( \hat{\beta} = BY \), which implies that \( BX = I \), a unit vector.

Hence

\[
\sum_{j=1}^{h} \lambda_j q_j^\prime E(\hat{\beta} - \beta)(\hat{\beta} - \beta)^\prime q_j = \sum_{j=1}^{h} \lambda_j q_j^\prime B \sum B^\prime q_j
\]

\[
= \sum_{j=1}^{h} \lambda_j q_j^\prime (X^\prime \sum^{-1} X)^{-1} q_j + \sum_{j=1}^{h} \lambda_j q_j^\prime [B - (X^\prime \sum^{-1} X)^{-1} X^\prime \sum^{-1}] \sum
\]

\[
[B - (X^\prime \sum^{-1} X)^{-1} X^\prime \sum^{-1}] q_i
\]

\[
\text{... ... (4.4.6)}
\]

Since \( BX = I \), then from (4.4.6)

\[
B = (X^\prime \sum^{-1} X)^{-1} X^\prime \sum^{-1} + [B - (X^\prime \sum^{-1} X)^{-1} X^\prime \sum^{-1}]
\]

This gives \( B = (X^\prime \sum^{-1} X)^{-1} X^\prime \sum^{-1} \)

Hence \( \hat{\beta} = (X^\prime \sum^{-1} X)^{-1} X^\prime \sum^{-1} Y \)

**Some Improved Estimators of \( \beta \)**

By adopting iterative algorithm proposed by Breiman, Tsur and Zemel (1992), one can improve the efficiency of these estimators. The implementation of above algorithm proceeds as follows.

**Step I:** Set \( \beta^{(0)} \), and initial value of the parameter vector by using (4.3.5) or (4.4.4), depending upon the distribution of disturbance term.

**Step II:** Fill the missing \( Y_i \) values with the estimators.
Step III: Calculate new \( \beta \) - estimators using

\[
\hat{\beta}^{(r+1)} = (X'X)^{-1} X'Y (\hat{\beta}^{(r)})
\]

(or)

\[
\hat{\beta}^{(r+1)} = (X' \sum_j^{-1} X_j)^{-1} X' \sum_j^{-1} Y (\hat{\beta}^{(r)})
\]

Step IV: Return to Step II unless the norm of the difference vector

\[ |\hat{\beta}^{(r+1)} - \hat{\beta}^{(r)}| \]

decreases below some predetermined convergence requirement.

Step V: Once the convergence criterion is satisfied, take the last value

of \( \beta \) as an improved estimator.

4.5. Confidence Intervals for Regression Coefficients

The common way of constructing confidence intervals for the parameters
of general regression model is by means of the standard errors of the
estimators. Since the estimators obtained by least-squares method are Best
Linear Unbiased Estimators (BLUEs), the confidence interval for \( \beta \) are

\[
\hat{\beta}_1 \pm Z_{\alpha/2} \text{SE} (\hat{\beta}_1)
\]

with \( Z_{\alpha/2} \) denote the upper \( \alpha/2 \) - th percentile of the Normal distribution.

Consider a simple regression equation of the form

\[
Y_j = \alpha_1 + \beta X_j + \epsilon_j
\]

The estimators for \( \beta \) and \( \alpha_1 \) are

\[
\hat{\beta} = \frac{\sum (X_j - \bar{X})^2 Y_j}{\sum (X_j - \bar{X})^2}
\]

and

\[
\hat{\alpha}_1 = Y - \hat{\beta} X
\]

The variances of \( \hat{\beta} \) and \( \hat{\alpha}_1 \) are
\[
\text{Var}(\hat{\beta}) = \frac{\sum (X_i - \bar{X})^2 \text{Var}(Y_i)}{\left(\sum (X_j - \bar{X})^2\right)^2} = \frac{\sigma^2}{\sum (X_i - \bar{X})^2}.
\]

\[
\text{Var}(\hat{\alpha_1}) = \text{Var}(Y) + \bar{X}^2 \text{Var}(\hat{\beta}) - 2 \bar{X} \text{Cov}(Y, \hat{\beta})
\]

\[
= \sigma^2 \left(\frac{1}{h} + \frac{\bar{X}^2}{\sum (X_j - \bar{X})^2}\right) \quad \text{since Cov}(Y, \hat{\beta}) = 0.
\]

The estimator of \(\sigma^2\) is

\[
s^2 = \frac{1}{(h-2)} \sum_{j=1}^{h} (Y_j - \hat{\alpha}_1 - \hat{\beta} X_j)^2.
\]

Hence the confidence intervals for \(\beta\) and \(\alpha_1\) are

\[
\hat{\beta} \pm s \ t_{h-2} (1 - \alpha/2) \left(\sum (X_i - \bar{X})^2\right)^{-1/2}
\]

and

\[
\hat{\alpha}_1 \pm s \ t_{h-2} (1 - \alpha/2) \left(\frac{1}{h} + \frac{\bar{X}^2}{\sum (X_j - \bar{X})^2}\right)^{1/2}.
\]

In case of general linear regression model, with \(\hat{\beta} = (X' X)^{-1} X' Y\), one can observe that \(E(\hat{\beta}) = \beta; \text{Var}(\hat{\beta}) = \sigma^2 (X' X)^{-1}\). Further, for any \(j\), \(\hat{\beta}_j\) is \(N(\beta_j, \alpha_{jj} \sigma^2)\), where \(\alpha_{jj}\) is the jth diagonal element in \((X' X)^{-1}\) matrix. The statistic

\[
t = \frac{\hat{\beta}_j - \beta_j}{\sqrt{\sum_{j} \frac{e^2_j}{h-k} \frac{1}{\sqrt{\alpha_{jj}}}}} \quad \text{has t distribution with h–k degrees of freedom.}
\]

It follows that the \(100(1 - \alpha/2)\) per cent confidence interval for \(\beta_j\) is

\[
\hat{\beta}_j \pm t_{h-k} (1 - \alpha/2) \sqrt{\sum_{j} \frac{e^2_j}{h-k} \frac{1}{\sqrt{\alpha_{jj}}}}.
\]
Venzon and Moolgavkar (1988) discussed the computation of profile-likelihood-based confidence intervals. They suggest that a computationally efficient method for the calculation of these confidence intervals consists of using the Newton-Raphson method for solving the system of non-linear equations

$$\frac{\partial L}{\partial \beta_i} = 0$$

Minkin (1992) calculated Profile-likelihood-based confidence intervals for $\beta_i$ in a general regression model. The justification of this procedure is the well-known result on the asymptotic distribution of likelihood estimators. The precision of the estimators are usually assessed by means of standard errors, with the confidence intervals constructed based on the asymptotic normality of maximum likelihood estimators.