Discussion Paper

CHUNG-HUA INSTITUTION FOR ECONOMIC RESEARCH
75 Chang Hsing St., Taipei, Taiwan
Republic of China

Discussion Paper

CHUNG-HUA INSTITUTION FOR ECONOMIC RESEARCH
75 Chang Hsing St., Taipei, Taiwan
Republic of China

DISCUSSION PAPER

No: 8402

AN EM ALGORITHM FOR THE HETEROSCEDASTIC REGRESSION MODELS WITH CENSORED DATA

Chihwa Kao

Chung-Hua Institution for Economic Research

This paper proposes an EM algorithm for the heteroscedastic regression models with censored data. The uniqueness of the EM algorithm is discussed. An iteratively reweighted least squares estimator is proposed.

AN EM ALGORITHM FOR THE HETEROSCEDASTIC REGRESSION MODELS WITH CENSORED DATA

Chihwa Kao

Chung-Hua Institution for Economic Research

1. Introduction

This paper is concerned with the estimation of the heteroscedastic regression models when the data is randomly right censored. Of course, all results extend to the case of left censored models by simply changing the signs of all variables. The uniqueness of the EM algorithm is discussed. An iteratively reweighted least squares estimator is proposed.

2. Model

Consider the following regression model

$$T_i = \beta' X_i + \varepsilon_i \ , \ i = 1, \ . \ . \ . \ , T,$$

Where the ε_i's are independent with unspecified distribution F, i,e., mean 0 and variance $\sigma_i^2 = \exp(\alpha' Z_i)$, β and X_i are $K \times 1$, α and Z_i are $P \times 1$.

Fore some reason, we are unable to observe $T_1, \ldots T_n$, and instead observe

$$Y_i = \min [T_i, C_i],$$

and

$$\delta_i = I [T_i \leq C_i] \ ,$$

Where C_i are the censored values and I[.] is the indicator function (i.e., $\delta_i = 1$ if $T_i \leq C_i$ and $\delta_i = 0$ if $T_i > C_i$).

Dempster, Laird & Rubin (1977) introduced the EM algorithm for iterative computation of maximum-likelihood estimates with incomplete data. Each iteration of the EM algorithm consists of two steps: the expectation step and the maximization step. First, let's discuss the expectation step. Follow Buekley and James (1979), we define

$$Y_i^* = Y_i \delta_i + E\,(T_i | T_i > Y_i)\,(1 - \delta_i).$$

Then we have the following lemma.

Lemma 1:

$E(Y_i^*) = \beta' X_i, \quad i = 1, \ldots, T.$

Proof: For example, see Miller (1981, p. 151).

For simplicity, assume F is normally distributed. Then the following result can be proved easily.

Lemma 2:

$$E(T_i | T_i > Y_i) = \beta' X_i + \sigma_i \frac{\phi\left(\frac{Y_i - \beta' X_i}{\sigma_i}\right)}{1 - \Phi\left(\frac{Y_i - \beta' X_i}{\sigma_i}\right)}, \qquad (1)$$

Where ϕ and Φ are the standard normal density and distribution function, respectively.

Clearly we cannot observe all of Y_i^*, we substitute estimates for $E(T_i | T_i > Y_i)$. We define those estimates as,

$$\bar{Y}_i = \hat{E}(T_i | T_i > Y_i)$$

$$= \hat{\beta}'X_i + \hat{\sigma}_i \frac{\phi\left(\frac{Y_i - \hat{\beta}'X_i}{\hat{\sigma}_i}\right)}{1 - \Phi\left(\frac{Y_i - \hat{\beta}'X_i}{\hat{\sigma}_i}\right)} .$$

Then we have the estimates of Y_i^* ,

$$\hat{Y}_i^* = Y_i \delta_i + \bar{Y}_i (1 - \delta_i). \qquad (2)$$

The maximization step takes the estimated complete data $\hat{Y}_i^*$ and estimates β and α by maximum likelihood as though the estimated complete data were the observed data. The log-likelihood function of the complete data is

$$\log L = -\frac{1}{2} \sum_i \alpha' Z_i - \frac{1}{2} \sum_i \exp(-\alpha Z_i)(\hat{Y}_i^* - \beta' X_i)^2. \qquad (3)$$

Therefore the EM algorithm is defined by cycling back and forth between (1)' and (3) till the estimates of β and α converge to the limiting values.

3. The uniqueness of the EM algorithm

It is known that multiple solutions occur frequently in practice for the EM algorithm. It would be interesting to know about the conditions in which the limiting value was guaranteed to be unique.

First we will show that the log-concavity in (3) is not guaranteed in fixed samples. It can be shown that the Hessian matrix of (3) is given by

$$H = \begin{pmatrix} \sum_i \exp(-\alpha' Z_i) X_i X_i' & \sum_i \exp(-\alpha' Z_i)(Y_i - \beta' X_i) X_i Z_i' \\ \sum_i \exp(-\alpha' Z_i)(Y_i - \beta' X_i) Z_i X_i' & \frac{1}{2}\sum_i \exp(-\alpha' Z_i)(Y_i - \beta' X_i)^2 Z_i Z_i' \end{pmatrix}$$

Where the Hessian is the negative of second derivatives matrix. We note that the Hessian above is not positive definite, since for n = 1, the determinant is negative. We are unable to prove or disprove the uniqueness of the likelihood function from the Hessian above. Because log-concavity is only a sufficient condition, not a necessary one, for the uniqueness.

In the next theorem, we will give a sufficient condition for the log-concavity in (3) in large samples.

Theorem 1:

$-\left(\frac{\partial^2 Q_T}{\partial\theta\ \partial\theta'} \right)$ is postive definite almost surely if

$$\exp(\alpha_0' \underline{Z}) - [(\beta_0' - \beta')\underline{X}]^2 > 0,$$

where

(a) $Q_T = \left(\frac{1}{T}\right) \log L$

(b) $\theta = (\beta', \alpha')$ and $\theta_0 = (\beta_0', \alpha_0')$, where θ_0 is a vector of true parameter,

and

(c) $\underline{X}$ and $\underline{Z}$ are drawn from G, where G is the limiting distribution of G_T and G_T is the joint empirical distribution of X_i and Z_i.

Proof:

Define

$Q_T = \left(\frac{1}{T}\right) \log L$. Q_T may be written as:

$$Q_T = -\frac{1}{2T}\sum_i \alpha' Z_i - \frac{1}{2}\sum_i (-\alpha_0' Z_i)\exp[(\alpha_0'-\alpha')Z_i]$$

$$\{(Y_i-\beta_0' X_i)^2 + [(\beta_0'-\beta')X_i]^2\} + \frac{2}{T}\sum_i (Y_i-\beta_0' X_i)(\beta_0'-\beta')X_i.$$

Under resonable conditions, it can be shown that (Kao, 1983),

(a) $\frac{1}{T}\sum_i \alpha' Z_i \longrightarrow E(\alpha'\underline{Z})$ uniformly,

(b) $\frac{1}{T}\sum_i \exp(-\alpha_0' Z_i)(Y_i-\beta_0' X_i)^2 \longrightarrow 1$ almost surely

and uniformly,

and

(c) $\frac{1}{T}\sum_i (Y_i-\beta_0' X_i)(\beta_0'-\beta')X_i \longrightarrow 0$ almost surely

and uniformly.

Therefore combining (a), (b) and (c), we have shown that Q_T converges almost surely and uniformly to Q given by,

$$Q = E\{-\frac{1}{2}[\alpha_0'\underline{Z} - (\alpha_0'-\alpha')\underline{Z}]$$

$$- \frac{1}{2}\exp(-\alpha_0'\underline{Z})[(\beta_0'-\beta')\underline{X}]^2[\exp(\alpha_0'-\alpha')\underline{Z}]$$

$$- \frac{1}{2}[\exp(\alpha_0'-\alpha')\underline{Z}]\}.$$

The first and second derivatives are,

$$\frac{\partial Q}{\partial \beta} = E\{\exp(-\alpha'\underline{Z})[(\beta_0'-\beta')\underline{X}]\underline{X}\},$$

$$\frac{\partial Q}{\partial \alpha} = E\{\frac{1}{2}\underline{Z} + \frac{1}{2}\exp(-\alpha'\underline{Z})[(\beta_0'-\beta')\underline{X}]^2\underline{Z}$$

$$+ \frac{1}{2}\exp[(\alpha_0'-\alpha')\underline{Z}]\underline{Z}\},$$

$$\frac{\partial^2 Q}{\partial \beta \partial \beta'} = -E[\exp(-\alpha'\underline{Z})\underline{X}\underline{X}'],$$

$$\frac{\partial^2 Q}{\partial \beta \partial \alpha'} = -E\{\exp(-\alpha'\underline{Z})\underline{X}\underline{Z}'[(\beta_0'-\beta')\underline{X}]\},$$

and

$$\frac{\partial^2 Q}{\partial\alpha\partial\alpha'} = -\frac{1}{2} E\{\exp(\alpha_0'\underline{Z})\exp(-\alpha'\underline{Z})\underline{Z}\underline{Z}' + \exp(-\alpha'\underline{Z}) [(\beta_0'-\beta')\underline{X}]^2\underline{Z}\underline{Z}'\}.$$

The matrix of the second derivatives can be written as,

$$-\left[\frac{\partial^2 Q}{\partial\theta\partial\theta'}\right] = -E\begin{pmatrix} a\underline{X}\underline{X}' & b\underline{X}\underline{Z}' \\ b\underline{Z}\underline{X}' & c\underline{Z}\underline{Z}' \end{pmatrix},$$

where

$$a = \exp(-\alpha'\underline{Z}),$$

$$b = \exp(-\alpha'\underline{Z})[(\beta_0'-\beta')\underline{X})],$$

and

$$c = \frac{1}{2}\exp(\alpha_0'\underline{Z})\exp(-\alpha'\underline{Z}) + \frac{1}{2}\exp(-\alpha'\underline{Z})[(\beta_0'-\beta')\underline{X}]^2.$$

Clearly, $-\left[\frac{\partial^2 Q_T}{\partial\theta\partial\theta'}\right] \longrightarrow -\left[\frac{\partial^2 Q}{\partial\theta\partial\theta'}\right]$ almost surely and uniformly.

For any K × 1 vector A and P × 1 vector B we have (e.g., Amemiya, 1973),

$$(A', B')\left[-\left(\frac{\partial^2 Q}{\partial\theta\partial\theta'}\right)\right]\begin{pmatrix} A \\ B \end{pmatrix}$$

$$= E[a(A'\underline{X})^2 + c(B'\underline{Z})^2 + 2b(A'\underline{X})(B'\underline{Z})]$$

$$\geq \lambda E[(A'\underline{X})^2 + (B'\underline{Z})^2],$$

where λ is the smallest eigenvalue of the following matrix M,

$$M = \begin{pmatrix} a & b \\ b & c \end{pmatrix}.$$

we note that λ is positive if M is positive definite. Therefore $\frac{\partial^2 Q}{\partial\theta\partial\theta'}$ is positive definite if M is positive definite. Next, we will give a sufficient condition for the positive definiteness of M.

Clearly, the determinant of M is given by

$$|M| = \frac{1}{2}\exp(-2\alpha'\underline{Z})\{\exp(\alpha_0'\underline{Z}) - [(\beta_0'-\beta')\underline{X}]^2\}.$$

Of course, the determinant of M is positive if

$$\exp(\alpha_0'\underline{Z}) - [(\beta_0'-\beta')\underline{X})^2] > 0.$$

Therefore λ is positive given the same condition above. Hence,

$-\left(\frac{\partial^2 Q_T}{\partial\theta\partial\theta'} \right)$ is positive definite almost surely if

$$\exp(\alpha_0'\underline{Z}) - [(\beta_0'-\beta')\underline{X}]^2 > 0. \qquad \text{Q.E.D.}$$

We note that $-\frac{\partial^2 Q_T}{\partial\theta\partial\theta'}$ will be positive definite almost surely every where if $\beta' = \beta_0'$, i.e., when β is at the true value β_0, or at the neighborhood of β_0, we also note that this result is nothing to do with the position of α.

4. An iteratively reweighted least squares estimator

In section 2 we note that the maximization step of the EM algorithm is to maximize (3), which is nonlinear in β and α. We also note that the uniqueness of the EM algorithm is not guaranteed. In this section we will propose a computationally simple method which the maximization step was based on Harvey (1976). The estimates of $\hat{\beta}_{k+1}$ and $\hat{\beta}_{k+1}$ at the (k+1)st step are

$$\hat{\beta}_{k+1} = [\sum_i \exp(-\hat{\alpha}_k'Z_i)X_iX_i']^{-1} \sum_i \exp(-\hat{\alpha}_k'Z_i)X_i\hat{Y}_i^*$$

and

$$\hat{\alpha}_{k+1} = [\sum_i Z_iZ_i']^{-1} \sum_i Z_i \log \hat{\mu}_i^2,$$

where

$\hat{\mu}_i = \hat{Y}_i^* - \hat{\beta}_k' X_i$ and $\hat{Y}_i^*$ is given in (2).

The iteration is continued until $\hat{\beta}_k$ and $\hat{\alpha}_k$ converge to the limiting value. With the possibility of non-uniqueness of the estimates of α and β, therefore, good starting values are important. For the choice of the starting values both for the method in section 2 and here, we propose

$$\hat{\beta}_0 = [\sum_i X_i X_i']^{-1} \sum_i X_i' Y_i$$

and

$$\hat{\alpha}_0 = [\sum_i Z_i Z_i]^{-1} \sum_i Z_i \log \mu_i^2, \quad i = 1, \ldots, T$$

where

$\hat{\mu}_i = Y_i - \hat{\beta}_0 X_i$.

5. Conclusion

In summary, this paper is an application of the EM algorithm to the heteroscedastic censored regression in a parameter form. This paper is also an alternative approach of the maximum likelihood estimate for the heteroscedastic censored regression. [(proposed by Kao (1983)]. Extensions to the nonparametric form of this model may be incorporated using the nonparametric product limit estimator of F based on the censored and uncensored residuals, i.e.,

$$F_\beta(\varepsilon) = 1 - \prod_{e_i \leq \varepsilon} \left(\frac{n-i}{n-i+1} \right)^{\delta_i}, \quad i = 1, \ldots, T,$$

where $e_i = Y_i - \hat{\beta}' X_i$ (Kaplan and Meier, 1958). The results of the nonparametric approach will be reported in another paper by the author.

References:

Amemiya, T., 1973, Regression analysis when the dependent variable is truncated normal, Econometrica 42, 997-1016.

Buckley, J. and James, I., 1979, Linear regression with censored data, Biometrika 66, 429-436.

Dempster, A.P., Laird, N.M. and Rubin, D.B., 1977, Maximum likelihood from incomplete data via the EM algorithm (with discussion), J.R. statist. Soc. B 39, 1-22.

Harvey, A.C., 1976, Estimating regression models with multiplicative heteroscedasticity, Econometrica 44, 461-464.

Kao, C.H., 1983, On the existence and uniqueness of the maximum likelihood estimate of censored normal regression with multiplicative heteroscedasticity, submitted to the Journal of Econometrics.

Kaplan, E.L. and Meier, P., 1958, Nonparametric estimation from incomplete observations, J. Am. Statist. Assoc. 53, 457-481.

Miller, R.G., 1981, Survival analysis (Wiley, New York).